

500.43115X00

THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): IWAMITSU, et al

Serial No.: 10/658,530

Filed: September 10, 2003

Title: A STORAGE SYSTEM AND A METHOD FOR DIAGNOSING FAILURE
OF THE STORAGE SYSTEM

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

September 26, 2003

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby
claim(s) the right of priority based on:

Japanese Patent Application No. 2003-195450
Filed: July 11, 2003

A certified copy of said Japanese Patent Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Carl I. Brundidge
Registration No. 29,621

CIB/rp
Attachment

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 7 月 1 1 日
Date of Application:

出 願 番 号 特 願 2 0 0 3 - 1 9 5 4 5 0
Application Number:

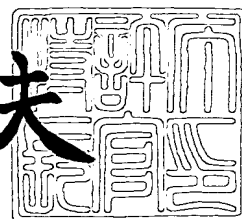
[ST. 10/C] : [J P 2 0 0 3 - 1 9 5 4 5 0]

出 願 人 株式会社日立製作所
Applicant(s):

2 0 0 3 年 8 月 2 6 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 3 - 3 0 6 9 6 5 0

【書類名】 特許願

【整理番号】 K03003071A

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

 【氏名】 岩満 幸治

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

 【氏名】 大鍋 賢二

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社日立製作所

【代理人】

 【識別番号】 100075096

 【弁理士】

 【氏名又は名称】 作田 康夫

【手数料の表示】

 【予納台帳番号】 013088

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶システム及び記憶システムの障害特定方法

【特許請求の範囲】

【請求項 1】 情報を格納する複数の記憶媒体と、
前記記憶媒体を制御するコントローラと、
前記記憶媒体および前記コントローラをループ状に接続し、前記コントローラおよび前記記憶媒体の相互間における情報の授受を行うループ状通信手段と、
前記コントローラは、障害が発生したループ状通信手段または、障害が発生した記憶媒体を認識した場合、上位装置から送られる情報の処理を行いながら、前記障害が発生したループ状通信手段または、前記障害が発生した記憶媒体を特定する処理を行うものであることを特徴とする記憶システム。

【請求項 2】 情報を格納する記憶媒体と、
前記記憶媒体を制御するコントローラと、
前記記憶媒体および前記コントローラをループ状に接続し、前記コントローラおよび前記記憶媒体の相互間における情報の授受を行うループ状通信手段と、
前記ループ状通信手段に信号の劣化を検出する信号劣化検出手段とを有するものであることを特徴とする記憶システム。

【請求項 3】 請求項 2 記載の記憶システムにおいて、
前記信号劣化検出手段が信号の劣化を検出し、信号が劣化したループ状通信手段で障害が発生した場合、前記コントローラは、前記信号劣化検出手段によって信号の劣化を検出した部位から障害の特定の処理を行うものであることを特徴とする記憶システム。

【請求項 4】 請求項 1 記載の記憶システムにおいて、
前記コントローラが記憶媒体に障害があることを特定できた場合、記憶媒体に障害があることを示す表示手段を有するものであることを特徴とする記憶システム。

【請求項 5】 請求項 2 記載の記憶システムにおいて、
前記コントローラが記憶媒体に障害があることを特定できた場合、記憶媒体に障害があることを示す表示手段を有するものであることを特徴とする記憶システム。

。

【請求項 6】 請求項 3 記載の記憶システムにおいて、
前記コントローラが記憶媒体に障害があることを特定できた場合、記憶媒体に障害があることを示す表示手段を有するものであることを特徴とする記憶システム

。

【請求項 7】 請求項 1 記載の記憶システムにおいて、
障害を特定できた段階毎に、障害部位を画面に表示する管理端末を有するものであることを特徴とする記憶システム。

【請求項 8】 請求項 2 記載の記憶システムにおいて、
障害を特定できた段階毎に、障害部位を画面に表示する管理端末を有するものであることを特徴とする記憶システム。

【請求項 9】 請求項 3 記載の記憶システムにおいて、
障害を特定できた段階毎に、障害部位を画面に表示する管理端末を有するものであることを特徴とする記憶システム。

【請求項 1 0】 請求項 1 記載の記憶システムにおいて、
前記コントローラは、前記障害が発生した部位を特定する処理を中断し、再開することができるものであることを特徴とする記憶システム。

【請求項 1 1】 請求項 2 記載の記憶システムにおいて、
前記コントローラは、前記障害が発生した部位を特定する処理を中断し、再開することができるものであることを特徴とする記憶システム。

【請求項 1 2】 請求項 4 乃至 6 記載の記憶システムにおいて、
前記表示手段は、LEDであることを特徴とする記憶システム。

【請求項 1 3】 請求項 3 記載の記憶システムにおいて、
前記コントローラは、前記障害が発生した部位を特定する処理を中断し、再開することができるものであることを特徴とする記憶システム。

【請求項 1 4】 情報を格納する記憶媒体と、前記記憶媒体を格納する筐体と、
前記記憶媒体を制御するコントローラと、前記コントローラに接続され、前記記憶媒体および前記コントローラをループ状に接続し、前記コントローラおよび前記記憶媒体の相互間における情報の授受を行うループ状通信手段とを有する記憶

システムであって、
前記コントローラは、障害が発生したループ状通信手段または、障害が発生した記憶媒体を認識した場合、上位装置から送られる情報の処理を行いながら、障害が発生した部位を特定するために筐体毎にループ状通信手段を切り離し、
前記切り離されたループ状通信手段に対して障害が発生した筐体の特定を行い、
障害が発生した筐体を特定できた場合、前記障害が発生した筐体内の記憶媒体またはループ状通信手段に対して障害の特定を行うことを特徴とする障害特定方法。

【請求項 15】 請求項 14 記載の記憶システムにおいて、
前記コントローラが記憶媒体に障害があることを特定できた場合、記憶媒体に障害があることを示す表示手段を点灯させることを特徴とする障害特定方法。

【請求項 16】 請求項 14 記載の記憶システムにおいて、
前記コントローラが記憶媒体に障害があることを特定できた場合、障害部位を画面に表示する管理端末を有することを特徴とする障害特定方法。

【請求項 17】 請求項 14 記載の障害特定方法において、
前記コントローラに所定以上の負荷がかかる場合、障害特定を一時的に中断し、
前記コントローラにかかる負荷が所定以下になる場合、再び障害特定を行うことを特徴とする障害特定方法。

【請求項 18】 請求項 15 記載の障害特定方法において、
前記表示手段は、LEDであることを特徴とする障害特定方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、記憶システムおよび情報処理技術に関し、特に、複数の構成要素をファイバチャネルループ等のループ状通信手段にて接続した構成の記憶システムおよび情報処理システム等に適用して有効な技術に関する。

【0002】

【従来の技術】

近年の情報処理技術の発展によって、情報処理システムおよび記憶システムの

役割が急速に大きくなってきている。情報処理システムおよび記憶システムに対する高性能化、高信頼性化や大容量化の要求は、高まる一方である。また、ネットワーク技術においても高速化の要求は、高まる一方である。

【0003】

このような超高速ギガビット・ネットワーク技術の一つとして、ファイバチャネル（FC）という技術が知られている。FCを使用することによって複数のハードディスクドライブを記憶システムのコントローラとをループ状に接続し、一つの記憶記憶システムを構築することが可能である。ファイバチャネル（FC）のループを用いた接続方式には、FC-AL（Fibre Channel-Arbitrated Loop）がある。記憶システムのコントローラとハードディスクドライブをループ状に接続する。

【0004】

FCループの規格上、FCループの一カ所でも切断や障害等が発生すると、コントローラとハードディスクドライブとの通信ができなくなるため、障害が発生したFCループ全体が使用不可能となる。

【0005】

そこで、ハードディスクドライブの交換や障害の場合等に対応するため、FCループの一部やハードディスクドライブをバイパス（切り離し）するポートバイパス回路（PBC：Port Bypass Circuit）が設けられている。

【0006】

コントローラは、PBCを切替えてループ全体が通信不可能とならないように、ループを制御する必要がある。

【0007】

【特許文献1】

特開 2001-222385号公報

【0008】

【発明が解決しようとする課題】

情報処理システムの中でも特に記憶システムは、無停止稼動を要求されるものであるため、保守作業に伴って記憶システムの処理を停止させることは、無停止

稼動という要求に反することになる。

【0009】

従って、本発明の目的は、記憶システムにおいて、通常処理を停止させることなく、障害部位を特定し、回復作業を行うことが可能な記憶システムを提供することにある。

【0010】

また、本発明の他の目的は、記憶システムにおいて、障害部位の特定の精度を向上させ、回復作業を迅速、簡便かつ的確に行うことが可能な記憶システムを提供することにある。

【0011】

また、本発明の他の目的は、記憶システムにおいて、信号の劣化を検出し、障害が発生した場合、信号の劣化が検出された部位から障害の特定処理を行うことによって、迅速、簡便、かつ的確に障害部位の特定を行うことが可能な記憶システムを提供することにある。

【0012】

【課題を解決するための手段】

上記目的を達成するために、本発明は、少なくとも一つの記憶媒体と、記憶媒体を制御する少なくとも一つのコントローラと、記憶媒体及びコントローラをループ状に接続し、コントローラと記憶媒体との通信を相互に行うループ状通信手段とを含み、上位装置からのリード・ライト等の処理を行いながら、障害部位を特定することができることを特徴とする。

【0013】

また、本発明は、障害の予測を行うために、ループ状通信手段に少なくとも一つの信号の劣化の検出を行う信号検出手段を有することを特徴とする。

【0014】

本発明によれば、通常処理を行っている状態で、障害部位を特定することができる。

【0015】

【発明の実施の形態】

以下、本発明の実施の形態を図面を用いて説明する。

【0016】

(第1の実施の形態)

図1は、本実施の形態の記憶システム150における装置概観図の一例である。ラックフレーム120は、基本筐体100とその他の増設筐体110を格納するものである。基本筐体100の内部には、ハードディスクドライブ130とコントローラ群140が格納されている。さらに、増設筐体110は、記憶システムの記憶容量を増大させる目的で利用されるものであり、複数のハードディスクドライブ130を有する。さらに記憶容量を増大させたい場合には、ラックフレーム及び増設筐体110を設けることも可能である。

【0017】

図2は、記憶システムに150における内部構成の概念図である。本実施の形態では、記憶システムにおけるループ状通信手段の一例としてFC-AL (Fibre Channel Arbitrated Loop) (以下、FCループという) を用いた場合の例について説明する。

【0018】

記憶システム150は、コントローラ群140とFCループ270～273と複数のハードディスクドライブ130とで構成される。コントローラ群140は、冗長化のために二つのコントローラ220、221を有している。コントローラ群140は、ハードディスクドライブ130等の情報を格納する管理情報と一時的に情報を格納するキャッシュメモリ230、231及び、上位装置210とコントローラ間の通信を行い、信号の変換を行うFCアダプタ240、241とコントローラ群140とハードディスクドライブ130との信号の変換等を行うFCアダプタに接続されている。ハードディスクドライブ130とFCアダプタ242、243との間には、PBC260～263とFCループ270～273に接続されている。コントローラ220、221は、FCループ270～273を制御する。

【0019】

2つのコントローラ間の通信は、コントローラ間通信線250によって行う。

コントローラ 220、221 は、上位装置 210 からのコマンドに基づいて、ハードディスクドライブ 130 の制御を行う。FC アダプタ 240、241 は、コントローラ 220、221 と上位装置 210 とでやりとりされる信号の変換等の処理を行う。FC アダプタ 242、243 は、コントローラ 220、221 とハードディスクドライブ 130 とで通信される信号の変換を行うと共に、PBC 260～263 の切替え処理も行う。PBC 260～263 は、交替パス 281～283 に接続することにより、コントローラ 220、221 が使用する FC ループを切替えることができる。例えば、PBC 260 は、FC ループ 270 から交替パス 280 に切替えることによって、FC ループ 272 と接続することができる。本実施の形態では、FC ループの切替えを行う手段の一例として PBC で行っているが、PBC に限らず、スイッチ等の他の手段を用いることもできる。

【0020】

FC ループ 270～273 は、基本筐体 100 と増設筐体 130 内に格納されている。FC ループ 270～273 と筐体との関係は、後述する。

【0021】

ハードディスクドライブ 130 は、簡略化のために、図 1 の複数のハードディスクドライブ 130 の台数により少なくしてある。複数のハードディスクドライブ 130 は、FC ループ 270、271 及び、FC ループ 272、273 に対して交互に接続されている。障害が発生していない正常な場合、PBC 260～263 は、それぞれ、FC ループ 270～273 に接続されており、コントローラ 220 は、FC ループ 270、271 を制御し、また、コントローラ 221 は、FC ループ 272、273 を制御する。

【0022】

上記の制御方法に適している一例として、たとえば、コントローラ 220、221 は、上位装置から送られるデータを分散して複数のハードディスクドライブ 130 に対して格納することにより、スループットの向上を実現するディスクアレイシステムが考えられる。ディスクアレイシステムにおいて、コントローラ 220、221 は、複数のハードディスクドライブ 130 に、上位装置との間で授受されるデータおよび、データから生成された冗長データを、論理ユニット 28

5～287のように分散して格納される。これにより、格納データの信頼性の向上が図れる。例えば、論理ユニット285において、一つのハードディスクドライブ130が故障した場合、正常なハードディスクドライブにあるデータと冗長データから故障したハードディスクにあったデータを復元することができる。さらに、本実施の形態において、2つのコントローラ220、221が同じ論理ユニットを制御すると、データの整合性を2つのコントローラ間で調整する必要があるため、データの処理速度が低下する。そこで、本実施の形態においては、コントローラごとに制御する論理ユニットが決められているものとする。どのコントローラがどの論理ユニットを制御するかについての情報は、キャッシュメモリ230、231に格納されている。

【0023】

FCループ270～273やハードディスクドライブ130に障害がある時、コントローラ220、221は、PBC260～263によって、交替パス280～283に接続させ、障害部位を含まないFCループとは異なる別のFCループ（以下、冗長ループという）を使用することができる。本実施の形態では、一例として、FCループを切り替える単位は、FCループ270、271または、FCループ272、273の2本単位とする。しかし、それぞれのFCループ毎にFCループを切り替えることも可能である。

【0024】

障害が発生した場合、コントローラ221に接続された管理端末の画面に、障害部位が表示される。管理端末により、FCループ270～273からハードディスクドライブ130の切り離しや、論理ユニットの設定の作業を行うことができる。本実施例において、管理端末280と記憶システム150は、管理端末用信号線281によって接続されている。管理端末用信号線281は、LANケーブル、RS-232Cや光ファイバー等を用いることができる。また、管理端末280の機能を上位装置210に移し、管理端末280を省略することもできる。

【0025】

図3は、図2のFCループの詳細図である。以下、図3と図1、2との対応関

係を説明する。上述の通り、FCループは、基本筐体と増設筐体を跨って構成され、さらに筐体内ループ、PBCと筐体間ループによって構成されている。図2のFCループ270は、筐体内のFCループ350、354、358及び、筐体間のPBC300、310、320、330、340及び、筐体内のPBC370、374、378にによって構成されている。同様にFCループ271は、図3のFCループ351、391、355と359に、FCループ272は、図3のFCループ352、392、356と360に対応する。

【0026】

一方、ハードディスクドライブ390～395は、筐体内のPBC370～381を介してFCループ350～361に接続されている。ハードディスクドライブ390、392、394は、FCループ270とFCループ272に接続されている。ハードディスクドライブ391、393、395は、FCループ271とFCループ273に接続されている。

【0027】

バイパス制御信号線1801は、各PBCとFCアダプタ242に接続されている。同様に、バイパス制御信号線1802は、各PBCとFCアダプタ243に接続されている。コントローラ220、221は、FCアダプタ242、243を経由し、バイパス制御信号線1801、1802に接続されたPBCにより、切り離し（バイパス）を行う。

【0028】

図4は、本実施の形態におけるPBCの構成の一例を示す。図4（a）のPBCのセクタ410は、各FCループに対するハードディスクドライブ、コントローラ又はFCループの一部分をバイパスする。LED420は、バイパス制御信号線430がON（バイパス状態）のときに点灯して外部にバイパス状態であることを報知する。バイパス制御信号線430は、記憶システム外部の制御端子等からも入力可能である。

【0029】

FCループまたは、ハードディスクドライブに障害が発生した場合、コントローラは、FCアダプタにバイパスの指示することによって、PBCにFCループ

をバイパスさせ、障害部位を切り離す。その後、コントローラは、障害部位を切り離した縮退運転をすることになるが、通常処理を継続させることができる。

【0030】

また、図4 (b) のPBCは、PBC260～263に使用され、ループの切替えをする。

【0031】

本実施の形態の記憶システムでは、一例として、FCループにおけるファイバチャネルプロトコルのデータリンク層として、SCSI-FCP (SCSI-3 Fibre Channel Protocol) を用いる。

【0032】

このSCSI-FCPを用いる場合、記憶システムはSCSI-FCPのイニシエータであるコントローラからターゲットとしてのハードディスクドライブの各々に対して発行されるFCPコマンドのSCSI Send Diagnosticsにて、パラメータ・リストで指定することにより、バイパス制御信号線430のON/OFFの出力指令を、ハードディスクドライブの各々に指示することが可能である。

【0033】

これにより、コントローラ220、221は、PBCのバイパス動作を制御する。また、同様に、現在のPBCにおけるバイパスの有無の状態は、FCPコマンドのSCSI Receive Diagnostic Resultsにて、ハードディスクドライブの各々を介して、コントローラが知ることができる。

【0034】

一方、図4に例示されるように、FCループの途中に設けられたPBCの各々は、FCループの途中の切り離し (バイパス) を行うセクタ410と、外部からこのセクタ410の切換え動作の制御のために入力されるバイパス制御信号線430と、バイパス状態のときに点灯し、外部にバイパス状態であることを報知するLED等の表示ランプ420等で構成されている。

【0035】

例えば、ハードディスクドライブ392をバイパスする場合を考える。コント

ローラ 2 2 0 は、バイパスするようにコマンドを発行する。コマンドを受領した F C アダプタ 2 4 2 は、バイパス制御信号線 1 8 0 1 を介して、P B C 3 7 4 のセレクト 4 1 0 により、ハードディスクドライブ 3 9 2 を F C ループ 3 5 4 からバイパスさせる。一方、L E D 4 2 0 は、バイパス状態であるので、点灯する。

【 0 0 3 6 】

なお、本実施の形態の記憶システムでは、F C ループとして、光ファイバや導線等の通信媒体を用いることに限らず、実装ボード上の配線パターンにて F C ループを構成することも含まれる。

【 0 0 3 7 】

以下、図面を使用して、記憶システムが上位装置からの読み込みや書き込みの通常処理を行いつつ、障害を特定する方法を説明する。

【 0 0 3 8 】

図 5 は、障害特定処理の全体のフローチャートである。F C ループ又はハードディスクドライブで障害が発生した場合（ステップ 5 0 1）、コントローラは、冗長ループに切替える（ステップ 5 0 2）。冗長ループに切替え終了後、再び、上位装置 2 1 0 からの読込や書き込みの処理を再開する（ステップ 5 0 3）。書き込み処理の場合、障害が記憶システムに発生していても、上位装置から書き込み処理を記憶システムが受け取り、キャッシュメモリに格納した段階でコントローラは、上位装置に、書き込み完了を通知する。

【 0 0 3 9 】

一方、読み込み処理で、記憶システムに障害が発生している場合、記憶システムが上位装置から読み込み処理を受け取った場合、冗長データを使用してデータを復元した後、障害ループから冗長ループに切替え、コントローラは、障害特定処理を実行できる条件を満たすかどうかの判断を行う。所定の条件を満たせば（ステップ 5 0 4）、コントローラは、障害特定処理の準備として、上位装置からの処理をキャッシュメモリに蓄積した後、冗長ループから障害が発生した F C ループに再び切替える（ステップ 5 0 5）。切替え完了後、コントローラは、障害が発生した筐体を調べるために、順番に筐体を切り離す。この操作を繰り返して障害が発生した障害筐体を調べる（ステップ 5 0 6）。

【0040】

コントローラは、障害が発生した障害筐体を特定できた段階で、管理端末に通知する。管理端末は、図示しない管理端末の表示画面に図2または図3の構成を表示する。コントローラは、この段階で障害特定処理を一時中断することも出来るし、このまま続けて障害特定処理を行うこともできる。事前に保守員は、障害特定処理を続けるかどうかの可否を設定しておいてもよいし、障害特定処理の続行をするかどうかについて所定の条件を設定して決めてもよい（ステップ507）。

【0041】

コントローラが障害特定を一時中断するのであれば、通常処理を行うために、障害ループから冗長ループに切替える（ステップ515）。冗長ループに切替えた後、上位装置からの処理を行う（ステップ516）。コントローラは、ステップ504と同様に障害特定処理を行う所定の条件を満たせば、再び障害特定処理を行う（ステップ517）。

【0042】

筐体毎の障害特定処理が終了し、筐体内の障害であると特定した場合、コントローラは、筐体内の障害特定処理を行う。同様に、障害特定処理を一時中断していた状態から再び障害特定処理を行う場合も、コントローラは、筐体内の障害特定処理を行う（ステップ509）。筐体内の障害では、ハードディスクドライブの障害又はFCループの障害が考えられる。一方、筐体間ループの障害であれば、障害特定処理を完了し（ステップ524）、通常処理を行う（ステップ525）。

【0043】

筐体内の障害特定処理は、各ハードディスクドライブに設けてあるPBCにより、FCループを切替えることによって、障害特定処理を行う。まず、コントローラは、筐体内のハードディスクに設けてられている全てのPBCにより、FCループを切り離す。その後、コントローラは、ハードディスクドライブに設けてあるPBCを一つだけ接続して、障害特定を行う。繰り返してこの処理を繰り返すと、障害が発生した部位に接続されたPBCを発見することができる。

【0044】

この段階で、ハードディスクドライブかまたは、PBCからハードディスクドライブまでのFCループに障害が発生していることになる。上記2つの部位のうち、一方に障害が発生していることになるため、コントローラは、ハードディスクドライブの障害であるかを確認するために、障害ループから冗長ループに切替える。コントローラは、障害が発生している可能性があるハードディスクドライブを冗長ループから調べることによって、ハードディスクドライブの障害であるかを確認することが出来る（ステップ511、512）。ハードディスクドライブの障害でない場合、PBCからハードディスクドライブまでのFCループに障害があることになる。

【0045】

筐体内の障害がこの段階まで特定できると、コントローラは、管理端末に障害部位を通知する。管理端末は、管理端末の画面に障害部位を表示する。さらに、コントローラは、障害特定処理を完了し、通常処理に戻る（ステップ514、528）。

【0046】

障害特定処理の一時中断は、本実施の形態において、筐体毎の障害特定処理の終了後と筐体内の障害特定処理の終了後に行うことにしたが、これに限定されることはない。例えば、筐体毎の障害特定処理で、一つの筐体に対して障害特定処理を終了した段階で、中断することもできる。あるいは、上位装置からの処理をキャッシュメモリに蓄積することが困難になった場合、障害特定処理を中断し、通常処理を行う。その後、再び障害特定処理を行うことができる。あるいは、コントローラは、上位装置で決められた記憶システムに対する応答時間内に障害特定処理が終了しない場合、コントローラは、障害特定処理を一時中断し上位装置の処理を再開する。その後、再び、コントローラは、障害特定処理を行う。

【0047】

図6から図11までは、図5の全体のフローチャートを詳細に説明したものである。ここで、障害が発生する部位は、FCループ270かFCループ270に接続されたハードディスクドライブであるとする。

【0048】

図6は、上位装置からの処理に対して、記憶システムが行う通常処理から障害が発生するまでの処理を説明したフローチャートである。上位装置から送られる書き込みまたは読み込みのコマンドは、上位装置とコントローラとの間に接続されたFCアダプタ242を介してコントローラ220へ送られる（ステップ604）。コントローラ220は、コマンドを解釈し書き込みのコマンドかまたは、読み込みのコマンド等を解釈する。コントローラは、コマンドが要求するデータ範囲に自コントローラが処理を行うものであるかを確認するため、キャッシュメモリ230を確認する。ここで、コントローラ220がキャッシュメモリ230を確認しているのは、処理を高速化するために、ハードディスクドライブに割り当てられた論理ユニットの処理を行うことができるコントローラが決められているからである。どのコントローラが、どの論理ユニットの処理を行うことができるかについては、予め決めてあり、キャッシュメモリ230、231に対応テーブルが格納されているものとする。

【0049】

書き込み処理で担当する論理ユニットが自コントローラの場合、コントローラ220は、キャッシュメモリ231に書き込み完了をした通知を受けた後、上位装置に書き込みが完了したことを通知する（ステップ606）。キャッシュメモリにある書き込みのデータを2重化のため、キャッシュメモリ231に書き込む。コントローラ220は、コントローラ間通信線250を使用して、書き込みデータをコントローラ221へ送信する。コントローラ221は、データを受領し、キャッシュメモリ231に書き込んだ後、コントローラ220へ書き込み完了を通知する。上位装置に書き込み完了を通知した後、ハードディスクドライブに書き込む処理を行う。ここで、上位装置からのデータをキャッシュメモリに2重に格納したのは、記憶システムの信頼性を向上させるためである。処理を高速に行う場合、コントローラ220は、もう一方のキャッシュメモリに書き込まない方法も考えられる。

【0050】

FCループに障害が起こっていない場合、データをハードディスクドライブ側

のFCアダプタを介してハードディスクドライブに書き込み処理を行う。

【0051】

しかし、コントローラがハードディスクドライブに書き込みのコマンドを発行して、所定の時間経過後、何の応答もない場合、再び書き込みのコマンドを再発行する（ステップ619）。コマンドの再発行回数が所定の回数を超過した場合、障害が発生したものと判定する（ステップ622）。コントローラ220は、障害が発生したものと判定したら、障害特定処理に移る。

【0052】

一方、読み込み処理は、以下のように行う。上位装置から読み込み処理を受け、コントローラ220がキャッシュメモリ230から読み込むべきデータがない場合、コントローラ220は、ハードディスクドライブからデータを読み込むように、読み込みのコマンドを発行する。書き込みの場合と同様に、所定の時間経過後、応答がない場合、タイムアウトとなる。コントローラは、再び読み込みのコマンドの再発行をし、コマンドの再発行回数が所定の回数を超過した場合、障害が発生したものと判定する。読み込み処理は、書き込み処理と異なり、上位装置にデータを送らないと読み込み処理が終了しない。コントローラ220は、障害が発生したものと判定した後、上位装置にデータを送るため他の正常なハードディスクドライブからデータを復元できるかを試みる。他の正常なハードディスクかデータを復元できない場合、FCループを冗長ループに切替える。冗長ループを使用しても、コントローラ220がデータを計算または読み込めない場合、上位装置へ読み込み失敗を通知する。その後、コントローラ220は、障害特定処理に移る。

【0053】

図7は、冗長ループを使用している記憶システムが上位装置からの読み込み書き込みの処理を行いながら、障害特定処理を開始するまでのフローチャートである。図5のステップ502からステップ504に対応する。コントローラ220は、障害が発生した障害ループから、冗長ループへ切替えるために、もう一方のコントローラ221にコントローラ間通信線250を使用して、冗長ループを使用することができるように指示する。指示を受け取ったコントローラ221は、現

在の処理を終了後、次の処理を、一時的にキャッシュメモリ 2 3 1 に蓄積させ、コントローラ 2 2 0 に処理完了を通知する（ステップ 7 0 4）。通知を受領したコントローラ 2 2 0 は、冗長ループを使用するために F C アダプタに対して F C ループ切替えの指示を出す（ステップ 7 1 6）。指示を受領した F C アダプタ 2 4 3 は、P B C 2 6 0、2 6 1 に交替パス 2 8 0、2 8 1 を使用して、F C ループ 2 7 2、F C ループ 2 7 3 に F C ループを切替えさせる。切替え完了後、F C アダプタ 2 4 3 は、コントローラ 2 2 0 へ切替え完了を通知する（ステップ 7 0 6）。コントローラ 2 2 0 は、切替え完了を受領後（ステップ 7 0 7）、コントローラ 2 2 1 に切替え完了を通知する（ステップ 7 0 8）。コントローラ 2 2 1 は、切替え完了の通知を受けた後（ステップ 7 0 9）、キャッシュメモリ 2 3 1 に蓄積されていた処理を再開する。その後、コントローラ 2 2 0 も、上位装置からの処理を再開する（ステップ 7 1 1）。この状態は、冗長ループを使用した縮退運転となる。このため、通信の帯域が 2 つの F C ループを使用していた場合に比べて半減し、性能が低下する。

【 0 0 5 4 】

縮退運転の間、コントローラ 2 2 0 は、障害部位を特定するため、所定の条件を満たすかを判定する。本実施の形態においては、コントローラ 2 2 0 が、コントローラ 2 2 0 とコントローラ 2 2 1 の両方の負荷が低く（ステップ 7 1 2）、キャッシュメモリ 2 3 0 と 2 3 1 の空き容量が十分にあり（ステップ 7 1 3）、かつ、過去の負荷状況の履歴から今後、負荷が高くなる可能性が低いと判定した場合（ステップ 7 1 4）、障害特定を開始する。尚、障害特定を開始する条件は、上記の条件に限られない。例えば、上記の条件で、何れか一つのみを満たす場合、障害特定を開始するようにしてもよい。

【 0 0 5 5 】

図 8 は、筐体毎の障害特定処理の開始準備から筐体毎の障害特定処理を行うフローチャートである。図 5 のステップ 5 0 5 とステップ 5 0 6 に対応する。障害特定処理は、上位装置からの書き込みや読み込み処理の影響を与えないように行われる。そのために、通常処理の途中に段階的に断続的に障害特定処理を行う。コントローラ 2 2 0 は、障害特定処理を行うために、もう一方のコントローラ 2

2 1 に障害特定処理の開始を通知する。コントローラ 2 2 1 は、通知を受領したら、上位装置からの現在の処理を終了させた後、次に行う処理をキャッシュメモリ 2 3 1 に蓄積する。ここで、障害ループコントローラ 2 2 1 の処理を一時中断させるのは、障害特定処理を迅速に行うためである。尚、コントローラ 2 2 1 を稼働させたまま、障害特定処理を行うこともできる。しかし、この場合、コントローラ 2 2 1 は、通常処理を行うことができるが、コントローラ 2 2 0 が障害特定処理を完了させる時間は、コントローラ 2 2 1 を停止させる方法に比べて、増大することになる。

【 0 0 5 6 】

コントローラ 2 2 0 は、コントローラ 2 2 1 の通常処理の停止を確認後（ステップ 8 0 7）、冗長ループから障害ループに切替えるために、F C アダプタ 2 4 3 にコマンドを発行する（ステップ 8 0 8）。コマンドを受領した F C アダプタ 2 4 3 は、P B C 2 6 0、2 6 1 に、冗長ループから障害ループに切替えさせる（ステップ 8 1 1）。

【 0 0 5 7 】

コントローラ 2 2 0 は、冗長ループから障害ループに切替え完了後、どの筐体で障害が発生しているかを確認するために、コントローラから一番遠い筐体を切り離して確認する。尚、本実施の形態では、一番遠い筐体から切り離すが、逆に一番近い筐体から切り離して、障害が発生した筐体を確認することも可能である。

【 0 0 5 8 】

コントローラ 2 2 0 は、一番遠い筐体をバイパスするように F C アダプタ 2 4 2 に指示を出す（ステップ 8 1 4）。F C アダプタ 2 4 2 は、P B C 3 3 0 によって、筐体をバイパスさせる。筐体のバイパス完了後、コントローラ 2 2 0 は、短縮した F C ループが正常であるかを確認するために、短縮した F C ループに障害特定コマンドを発行する（ステップ 8 1 9）。障害特定コマンドの一例として、ファイバチャネルの LIP (Loop Initialization Primitive) を使用することができる。また、障害特定コマンドも通常時のコマンドと同様に、リトライ回数とタイムアウト時間を設定することができる。障害特定コマンドを発行した後、F C

ループから応答があれば、前回バイパスした筐体または筐体間 F C ループに障害が発生していることになる。

【0059】

一方、障害特定コマンドを発行しても、短縮した F C ループから応答がない場合、F C ループをさらに短縮させる必要がある。F C ループをさらに短縮させるために、コントローラ 220 は、F C アダプタ 242 に F C ループを短縮させるようにコマンドを発行する。F C アダプタ 242 は、P B C 320 によって、筐体間 F C ループをバイパスする。この後、コントローラ 220 は、障害特定コマンドを発行することによって、応答を確認する。コントローラ 220 は、応答があるまで F C ループを短縮させる処理を行う。コントローラ 220 は、応答を受領後、さらに、障害特定処理を続行するかまたは、ここで、上位装置からの処理がタイムアウトにならないように障害特定処理を一時中断することになる。

【0060】

図 9 は、筐体内の障害特定処理を行うフローチャートである。図 5 のステップ 509 に対応する。障害が発生した部位が筐体間ループか筐体そのものであるかによって処理が異なる（ステップ 901）。筐体間ループに障害がある場合、コントローラ 220 は、さらに障害特定処理を行う必要がない。コントローラ 220 は、筐体間ループに障害が発生した障害ループが使用可能であるかを判断する（ステップ 903）。コントローラ 220 から障害が発生した部位より遠い F C ループに接続されたハードディスクドライブを使用していない場合、コントローラ 220 は、冗長ループを使用せずに、障害が発生した筐体間ループを切り離すことによって、障害ループを使用することができる。コントローラ 220 が障害ループを使用することができない場合、コントローラ 220 は、冗長ループを使用する（ステップ 904）。

【0061】

ステップ 901 に戻って、障害が発生した部位が筐体内である場合、コントローラ 220 は、筐体内の障害を特定するために、障害が発生した筐体内の全ての接続されたハードディスクドライブを P B C に切り離させる。この状態で、コントローラ 220 は、障害特定コマンドを発行する（ステップ 922）。障害特定

コマンドに対する応答がない場合、障害部位は、筐体内の F C ループに障害があることになる（ステップ 934）。一方、コントローラ 220 は、全てのハードディスクドライブに接続される P B C をバイパスさせ、障害がない場合、障害が発生した部位は、ハードディスクドライブになる。

【0062】

コントローラ 220 は、筐体内の P B C により、ハードディスクドライブを順番に一つずつ接続させる（ステップ 918）。障害部位に接続した P B C がある場合、コントローラ 220 が発行した障害特定コマンドの応答がなければ、障害部位は、ハードディスクドライブまたは、P B C とハードディスクドライブの間にある F C ループに障害があることになる。

【0063】

図 11 は、ハードディスクドライブで障害が発生しているものと判明した場合に、コントローラが行う処理についてのフローチャートである。図 5 の 512 から 514 に対応する。ハードディスクドライブに障害があるかどうかを確認するために、コントローラ 220 は、障害ループから冗長ループに切替える（ステップ 1101）。コントローラ 220 は、冗長ループに切替えた後、障害特定コマンドを発行する（ステップ 1106）。冗長ループで障害特定コマンドの応答があれば、P B C とハードディスク間のループに障害が発生していると特定することができる（ステップ 1114）。その後、障害部位をバイパスして障害ループを使用する方法と冗長ループを使用して、通常処理を行なうことになる。

【0064】

一方、ステップ 1110 で、冗長ループで障害特定コマンドの応答がなければ、ハードディスクドライブに障害が発生していることになる。ハードディスクドライブの故障であれば、コントローラ 220 は、障害が発生したハードディスクドライブの表示ランプ 160 を点灯させて、障害であることを保守員に知らせる（ステップ 1123）。障害が発生したハードディスクドライブの表示ランプ 160 を点灯させることによって、保守員は、簡便に障害が発生したハードディスクドライブを発見することができる。また、管理端末 280 の画面に障害が発生したことを表示する（ステップ 1120）。

【0065】

障害特定処理が終了すれば、通常処理に戻るための処理を行う。コントローラ 220 は、障害特定処理を終了させるために、コントローラ 221 に障害特定処理の完了を通知する（ステップ 1132）。完了の通知を受けたコントローラ 221 は、キャッシュメモリ 231 に蓄積された処理を再開する（ステップ 1134）。コントローラ 221 の通常処理再開後、コントローラ 220 も処理を再開する（ステップ 1135）。コントローラ 220 が障害特定処理を段階毎に中断する場合も障害特定処理の終了の処理と同様にコントローラ 221 に障害特定処理完了を通知する。コントローラ 220 は、コントローラ 220 の通常処理の開始を確認後、通常処理に移る。但し、中断処理の場合は、再び障害特定処理を開始する必要がある。コントローラ 220 は、コントローラ 221 との互いの負荷と、キャッシュメモリ 230 とキャッシュメモリ 231 との容量と将来の負荷状況を監視し、所定の条件で再び障害特定処理を開始する。

【0066】

コントローラ 220 が障害特定処理を段階的に行うことで、1 回あたりに必要とする障害特定処理の時間を大幅に短縮することができ、上位装置からの通常処理の影響を与えることなく障害特定処理を行うことができる。また、管理端末に障害部位の詳細を通知することで、保守員は、迅速かつ正確に障害部位の交換することができる。

【0067】

図 10 は、障害特定処理を一時中断し、再び障害特定処理を行うまでの処理である。コントローラ 220 は、障害特定処理をステップ 823 やステップ 825 等で中断することができる。または、上位装置からの読み込みや書き込み等の処理が上位装置の制限時間内に上位装置に対して応答できない場合、コントローラ 220 は、障害特定処理を一時中断することができる。図 10 は、図 5 のステップ 515 からステップ 518 に対応する。コントローラ 220 は、障害特定処理の中断箇所をキャッシュメモリに格納する。キャッシュメモリに格納された中断箇所の情報は、再び障害回復処理を行う際、使用する。コントローラ 220 は、キャッシュメモリに格納終了後、コントローラ 220 が、障害ループを引き続き

使用可能であるなら、コントローラ 220 は、その障害した F C ループを使用する（ステップ 1000）。コントローラ 220 は、コントローラ 221 に通常処理再開を通知した後（ステップ 1001）、コントローラ 220 とコントローラ 221 は、通常処理を行う。その後、コントローラ 220 は、障害特定処理を行う所定の条件を満たせば、障害特定処理を再開する（ステップ 1005～1008）。コントローラ 220 は、障害特定処理再開するために、コントローラ 221 に障害特定再開を通知する（ステップ 1014）。通知を受けたコントローラ 221 は、現在の処理を終了後、上位装置からの処理をキャッシュメモリに蓄積する（ステップ 1017）。その後、コントローラ 220 は、キャッシュメモリから障害特定処理の中断した情報を読み出し、障害特定処理を再開する。

【0068】

本実施の形態において、バイパス制御信号線 1801、1802 を設けて、コントローラがハードディスクドライブや F C ループをバイパスしたが、バイパス制御信号線を F C ループに含ませることも可能である。

【0069】

（第 2 の実施の形態）

障害が発生する前に F C ループの信号の劣化を監視する回路として信号劣化検出回路 1104 を F C ループ中に設けることができる。図 12 は、信号劣化検出回路 1104 の構成図である。信号物理検出部 1102 は、F C ループ中の物理的な信号の振幅を監視する。信号論理検出部 1101 は、信号の論理的な形式を監視する。ファイバチャネルプロトコルの一例として、信号論理検出部 1101 は、シーケンスやフレームの異常を検出することができる。信号物理検出部 1102 または信号論理検出部 1101 は、信号の劣化が所定の条件より悪化すると、劣化検出制御部 1103 に通知する。劣化検出制御部 1103 は、信号の劣化を F C ループを使用してコントローラに通知する。または、劣化検出制御部 1103 は、後述する信号線を使用してもよい。

【0070】

図 13 は、図 12 の信号劣化検出回路を筐体毎に記憶システムに組み込んだ場合における記憶システムの一部の構成図である。信号劣化検出回路 1201～

1206は、各筐体に2つずつ設けてある。本実施の形態では、一つの筐体に2つの信号劣化検出回路を設けたが、FCループのそれぞれの出入り口に信号劣化検出回路を設け信号劣化検出回路は、筐体間PBCの回路とともに設けることもできる。本実施の形態では、一例としてFCループのような単方向信号の場合について説明する。

【0071】

図14は、図13における信号劣化検出回路が信号の劣化を検出し、コントローラが障害特定処理を行うまでのフローチャートである。信号劣化検出回路が信号の劣化を検出した場合、コントローラに劣化を通知する（ステップ1301）。コントローラ220は、信号劣化検出回路から送られた信号劣化情報をキャッシュメモリに格納する。FCループの場合、単方向信号であるので、FCループの上流で障害が発生した場合、その障害が発生したFCループより下流のFCループの信号は、正常な信号とならない。よって、信号の劣化が発生した箇所は、正常な信号であるとした信号劣化検出回路と異常であると検出した信号劣化検出回路との間となる（ステップ1304）。ここで、コントローラ220に情報格納は図15のようにキャッシュメモリに格納される情報は、信号劣化検出回路番号と信号劣化検出回路が設けられている場所と信号の状態コントローラ220は、その信号劣化の発生した箇所を管理端末へ通知する（ステップ1305）。通知を受けた管理端末は、管理端末の画面に信号が劣化した箇所を表示し（ステップ1308）、保守員は、画面に表示された箇所を点検することができる。また、保守員は、故障が発生する前に信号を劣化させた部品を交換することができる。

【0072】

仮に、図13において、信号劣化検出回路1203は、正常であるが、信号劣化検出回路1204が信号の劣化を検出した場合、信号を劣化させる箇所は、信号劣化検出回路1204より上流であり、かつ、信号劣化検出回路1203より、下流にあることになる。さらに、信号劣化検出回路1203、1204は、筐体のFCループの出入り口に設けられているので、信号を劣化させた箇所は、筐体内にあることになる。

【0073】

また、信号劣化検出回路1202は、正常であるが、信号劣化検出回路1203が信号の劣化を検出した場合、信号劣化検出回路1203より上流であり、かつ、信号劣化検出回路1202より下流にあることになる。よって、信号を劣化させた箇所は、筐体間のFCループにあることになる。

【0074】

信号を劣化させた箇所は、障害になる可能性が高い。そこで、信号を劣化させたFCループで、障害が発生した場合、信号を劣化させた筐体から障害特定処理を行うことで、障害特定処理に要する時間を短縮することができる。

【0075】

図15は、信号を劣化させた筐体から障害特定処理を行う概略を示したフローチャートである。最初に、コントローラは、信号を劣化させた筐体内部から障害特定処理を行う（ステップ1400）。ここで、障害部位を特定できれば、終了する。しかし、コントローラが信号を劣化させた筐体内部の障害特定処理を行っても、障害部位を特定できない場合、他の筐体に障害が発生していることになる。この場合、順番に筐体を調べることになる（ステップ1402）。ステップ1402で、信号を劣化させた筐体は、一度、調べているので、障害特定処理を行う必要はない。コントローラが別の筐体で障害を特定できた場合、さらに、障害が発生した筐体内部を調べる（ステップ1403）。コントローラは、障害が発生した筐体内部で障害部位を特定できれば、障害特定処理を終了する。

【0076】

本実施の形態において、信号劣化検出回路が各筐体に設けられているので、障害が発生する前に信号の劣化を検出することができる。FCループに障害が発生した場合、コントローラは、予め、信号を劣化させている筐体を把握しているため、信号を劣化させた筐体から障害特定処理を行うことができ、短時間で障害特定処理を終了させることができる。

【0077】

（第3の実施の形態）

図13の変形例では、筐体毎ではなく、ハードディスクドライブ毎に信号劣化検出回路を設けた記憶システムである。本実施の形態を図示すると、図16になる。図16で、信号劣化検出回路は、ハードディスクドライブから出力される信号の劣化検出を行う。信号劣化検出回路が各ハードディスクドライブからの出力信号をチェックすることにより、筐体毎に信号劣化検出回路を設けた場合と比較して、コントローラは、信号を劣化させた部位について詳細に知ることができる。信号を劣化させた部位の精度は、信号劣化検出回路の間であるので、ハードディスクドライブかまたはFCループで信号を劣化させているかまで推定することができる。

【0078】

信号を劣化させた部位の特定する流れは、図14とほぼ同じである。異なる点は、ステップ1304で信号劣化箇所の推定を筐体毎ではなく、ハードディスク毎で行うことである。

【0079】

さらに、障害が信号の劣化を検出したFCループで発生した場合、処理の流れは、図17のようになる。障害が発生したFCループで、信号を劣化させた箇所で障害が発生している可能性が高いので、信号を劣化させた部分について障害が発生していないか障害特定処理を行う（ステップ1600）。この処理では、信号を劣化させた部位である一部のFCループとハードディスクドライブの障害特定処理を行う。まず、FCループに障害が発生しているかを確認するために、PBCによりハードディスクドライブとPBCからハードディスクドライブまでのFCループを切り離し、障害が発生しているかを確認する。次に、ハードディスクドライブで障害が発生しているかを特定するために、反対側のコントローラから信号を劣化させた可能性があるハードディスクドライブについて障害特定処理を行う。

【0080】

障害を発生させた部位を特定できない場合、同一筐体内の他の部位について障害特定処理を行う（ステップ1602）。コントローラは、同一の筐体内で障害を部位を特定できない場合、他の筐体について障害特定処理を行う（ステップ1

604)。

【0081】

ハードディスクドライブの信号が出力される部分に信号劣化検出回路を設けることによって、筐体に信号劣化検出回路を設けた場合と異なり、信号を劣化させた部位を詳細に知ることができる。

【0082】

(第4の実施の形態)

図18は、図13、16の変形例として信号劣化検出回路をハードディスクドライブの信号が入力される箇所と信号が出力される箇所にそれぞれ設けた記憶システムの変形例の一例である。本実施の形態における信号劣化検出回路は、ハードディスクドライブに入力される信号及びハードディスクドライブから出力される信号の両方について信号の劣化を検出することができるので、ハードディスクドライブの片方に設けた場合と比較して、信号を劣化させた部位をさらに特定することができる。特に、ハードディスクドライブが信号を劣化させた部位であるかを判定できる可能性が高い。

【0083】

信号が劣化したFCループで障害が発生した場合、信号を劣化させた部位で障害が発生した可能性が高いため、コントローラは、信号を劣化させた部位について最初に障害特定処理を行う。処理の流れは、図17とほぼ同じであるが、異なる点は、図17のステップ1600の処理である。信号を劣化させた部位がハードディスクドライブである場合、コントローラは、冗長ループを使用してハードディスクドライブの障害特定処理を行う。また、ハードディスクドライブ間のFCループで信号が劣化している場合、障害ループを使用してその信号を劣化させたFCループに対して障害特定処理を行う。本実施の形態では、ハードディスクドライブから入力される信号と出力される信号箇所のそれぞれに信号劣化検出回路を設けることによって、FCループが信号を劣化させているのかハードディスクドライブで信号を劣化させているのかを判定することができる。

【0084】

信号劣化検出回路を設けた記憶システムでは、FCループで信号の劣化を検出

した段階で、信号を劣化させた部位を特定し、障害が発生する前に保守員等に知らせることができる。さらに、信号を劣化させた部位が接続されているFCループに障害が発生したら、コントローラは、信号を劣化させた部位から障害特定処理を行うことによって、障害部位の発見を迅速にでき、性能や信頼性の低下を最小限に抑えることができる。さらに、障害の回復操作等の保守管理作業を迅速かつ的確に行うことが可能になる。

【0085】

(第5の実施の形態)

図19は、本実施の形態の概念図である。正常な場合、コントローラ1910は、FCループ1901～1904を制御し、コントローラ1911は、FCループ1905～1908を制御する。奇数筐体1912は、FCループ1901、1902と、FCループ1905、1906に接続されている。また、偶数筐体1913は、FCループ1903、1904と、FCループ1907、1908に接続されている。

【0086】

偶数筐体1913と奇数筐体1912は、各々使用するFCループが異なるため、隣接する筐体に影響を与えない。これにより、偶数筐体1913と奇数筐体1912は、別々に障害特定処理を行うことができ、信頼性を高めることができる。また、図2、3と比較してFCループを多重に設けていることからコントローラは、ハードディスクドライブと高速に読み込みや書き込みの処理をすることができる。

【0087】

以上、本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で変更可能であることはいうまでもない。

【0088】

たとえば、障害特定処理において、上述の説明では障害ありと特定した筐体内のハードディスクドライブのみ切り離して障害特定処理を行う。しかし、この方法だけに限らない。例えば、コントローラは、障害が発生したFCループに接続

されている全てのハードディスクドライブを全て切り離した後、障害が発生した筐体のハードディスクドライブを一台毎に障害の発生を調べる方法も考えれる。あるいは、障害が発生している筐体が複数ある場合、複数の障害が発生している筐体のハードディスクを切り離した後、障害を特定すべき筐体のハードディスクドライブを一台毎に接続して障害特定処理を行う。上記の方法においては、障害が発生したハードディスクドライブが2台以上であっても障害を回復することができる。

【0089】

また、障害箇所の表示方法としては、表示ランプ等を用いる方法に限らず、例えば、コントローラを外部から制御する管理端末280等の画面に図3のようにシステム構成図を表示し、このシステム構成図上に可視化して表示する方法でもよい。

【0090】

また、通信手段は、FCループだけに限られず、その他、光ファイバや銅線等のループ状の通信手段であれば、応用可能である。

【0091】

また、記憶手段は、ハードディスクドライブだけに限られず、光ディスクやメモリにも応用可能である。

【0092】

また、障害がある一定の確率でしか起こらず、再現性が困難な場合、コントローラは、障害とみなす確率を設定することができる。これにより、障害特定処理においても、コントローラは、障害特定コマンドを複数回発行し、応答した回数で応答率を算出し、所定の応答率以下なら、障害が発生しているとすることができる。

【0093】

【発明の効果】

本発明によれば、上位装置からの書き込みまたは読み込みなどの通常処理に影響を与えず、上位装置のタイムアウト時間内で記憶システムの障害特定処理を行うという効果が得られる。

【0094】

また、本発明によれば、ループ状通信手段を備えた構成において障害発生時の性能や信頼性の低下を最小限に止めることができる、という効果が得られる。

【0095】

本発明によれば、ループ状通信手段を備えた構成において、障害部位の特定、回復作業を迅速、簡便かつ的確に行うことができる、という効果が得られる。

【0096】

本発明によれば、ループ状通信手段を多重に備えた構成において、複数のループ状通信手段に及ぶ多重障害発生時の復旧を確実に行うことができる、という効果が得られる。

【図面の簡単な説明】

【図1】 第1の実施の形態における記憶システムの外観図である。

【図2】 第1の実施の形態における記憶システムのループ構成の概念図である。

。

【図3】 図2のループ構成の概念図における詳細な構成図である。

【図4】 PBCの構成図である。

【図5】 障害特定処理の全体フローチャートである。

【図6】 図5のフローチャートの一部である。

【図7】 図5のフローチャートの続きである。

【図8】 図5のフローチャートの続きである。

【図9】 図5のフローチャートの続きである。

【図10】 図5のフローチャートの続きである。

【図11】 図5のフローチャートの続きである。

【図12】 信号劣化検出回路の構成図である。

【図13】 信号劣化検出回路を記憶システムの筐体毎に組み込んだ構成図である。

【図14】 信号の劣化の検出から障害特定処理を行うまでのフローチャートである。

【図15】 信号を劣化させた筐体から障害特定処理を行う概略を示したフロー

チャートである。

【図 1 6】 信号劣化検出回路を記憶システムのハードディスクドライブの出口に設けた構成図である。

【図 1 7】 信号劣化検出回路を設けた場合の障害特定処理のフローチャートである。

【図 1 8】 信号劣化検出回路を記憶システムのハードディスクドライブの出入口に設けた構成図である。

【図 1 9】 奇数筐体と偶数筐体に分けた場合の記憶システムの構成図である。

【符号の説明】

基本筐体… 1 0 0、

増設筐体… 1 1 0、

ハードディスクドライブ… 1 3 0、

表示ランプ… 1 6 0、

上位装置… 2 1 0、

コントローラ… 2 2 0、2 2 1、

キャッシュメモリ… 2 3 0、2 3 1、

FC ループ… 2 7 0 ~ 2 7 3、3 5 0 ~ 3 6 1、3 9 0 ~ 3 9 7、

管理端末… 2 8 0

PBC… 2 6 0 ~ 2 6 3、3 0 0 ~ 3 4 0、3 7 0 ~ 3 8 1、

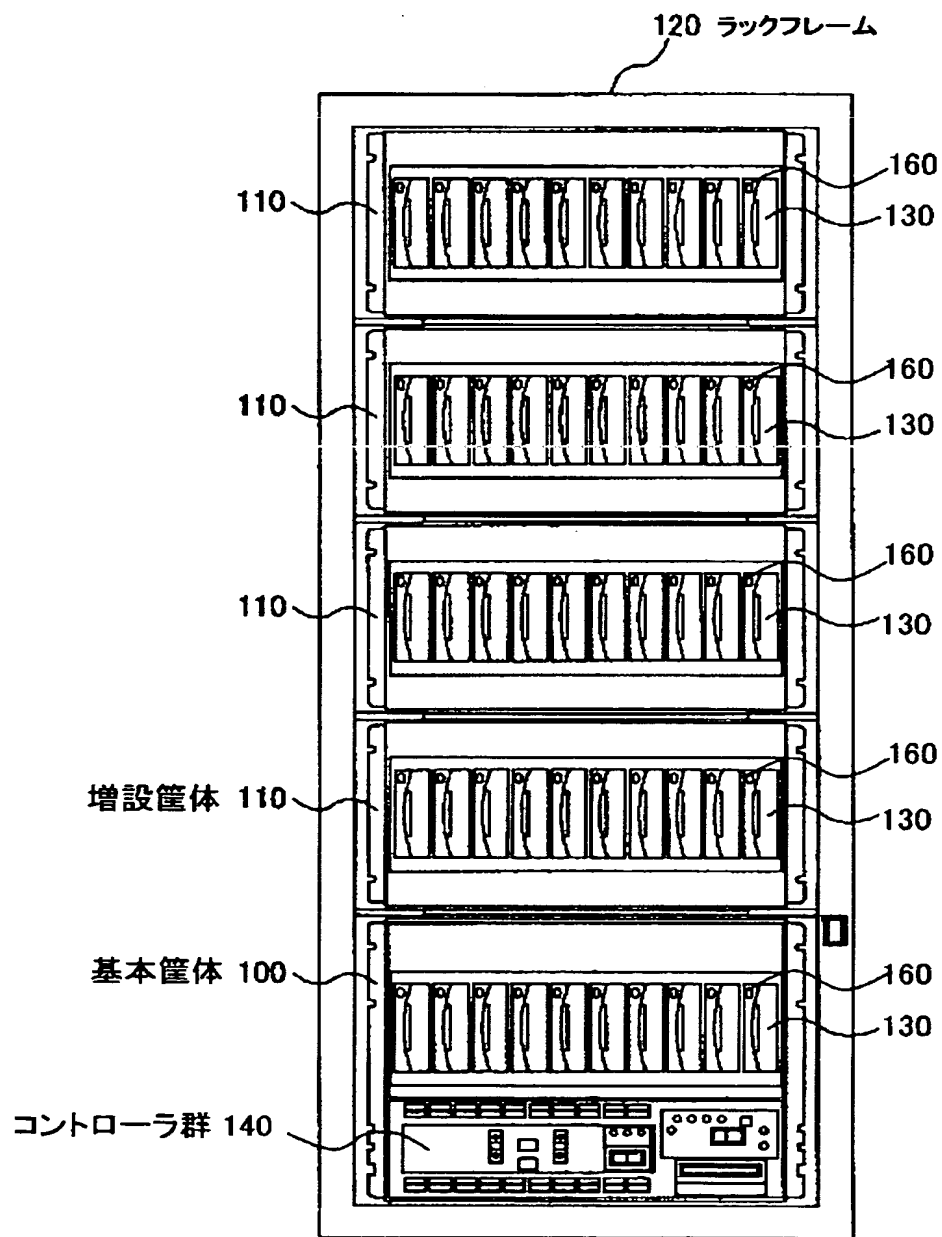
信号劣化検出回路… 1 1 0 4、

信号線… 1 8 0 1、1 8 0 2

【書類名】 図面

【図 1】

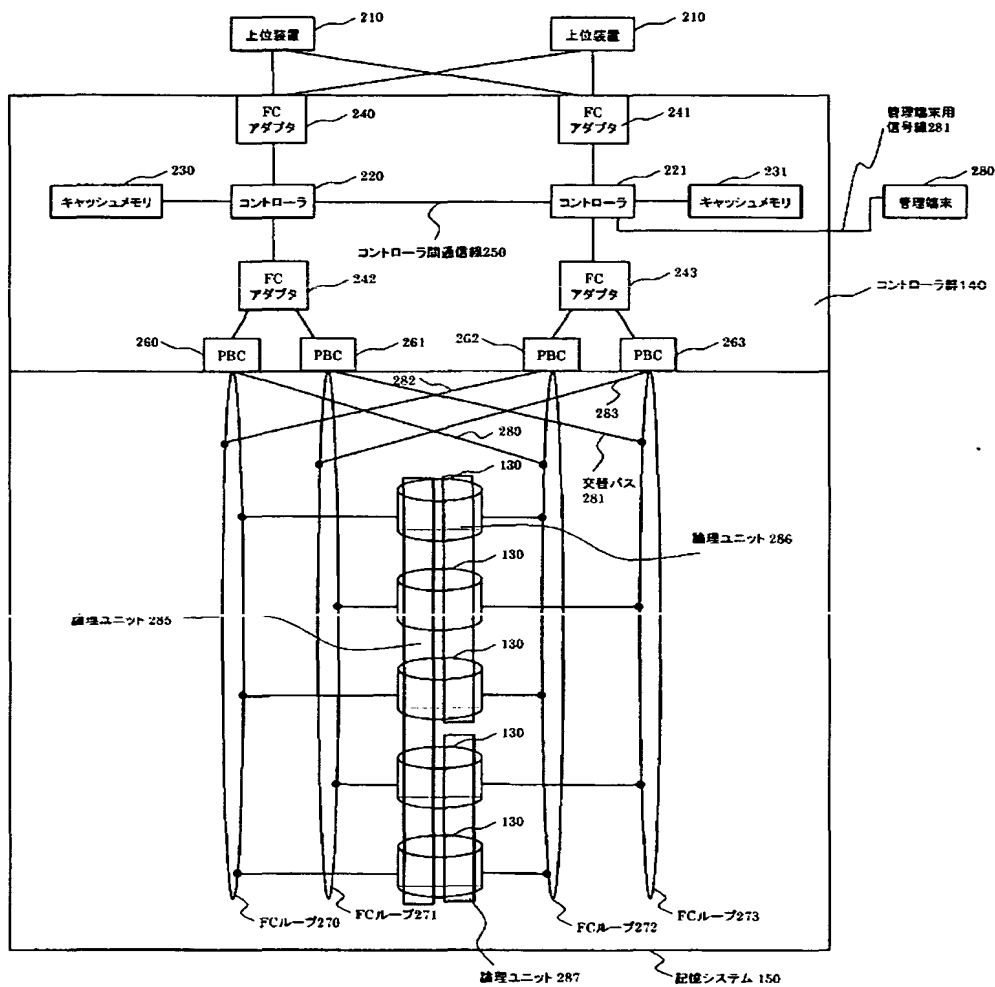
【図1】



記憶システム 150

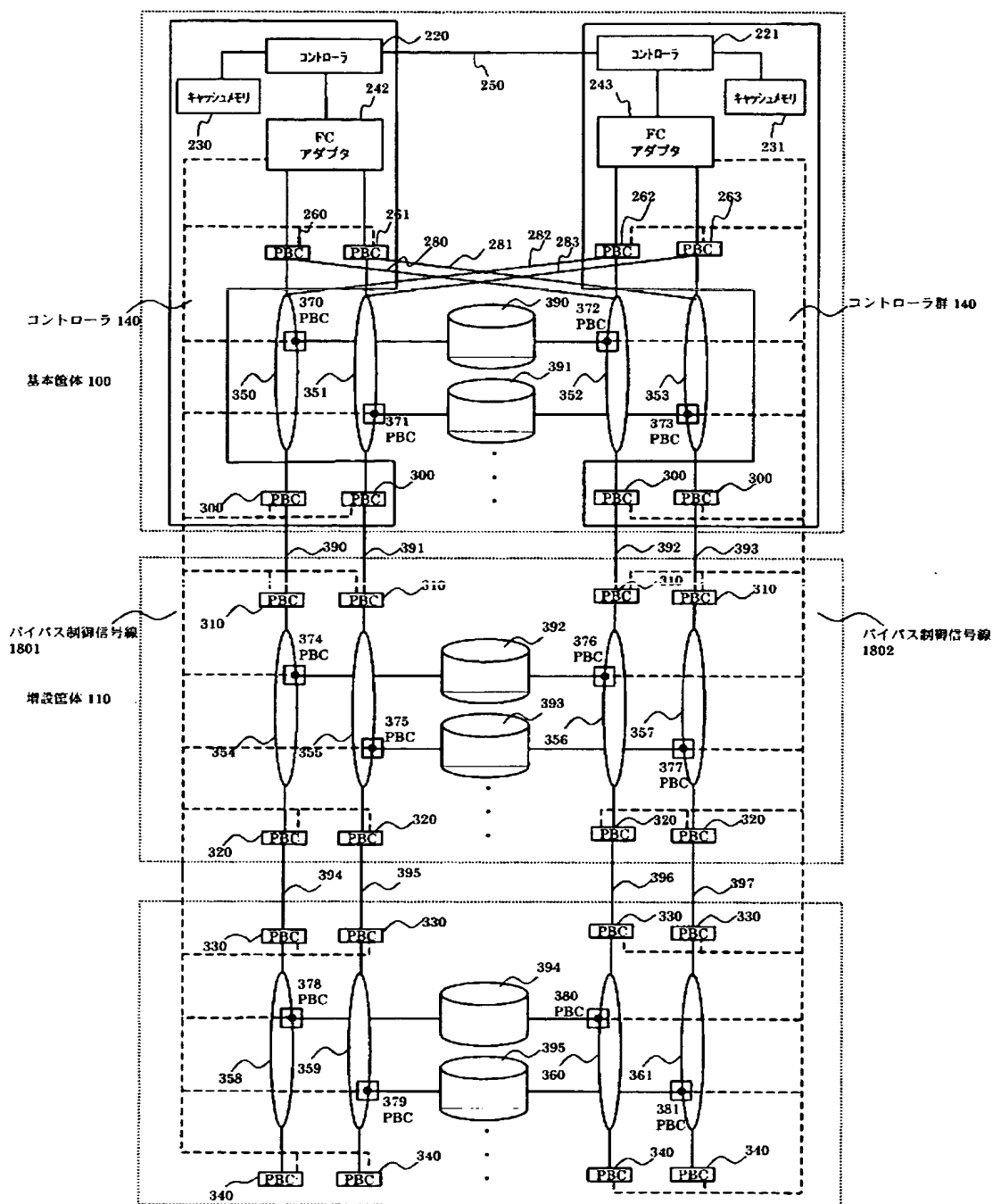
【図 2】

【図2】



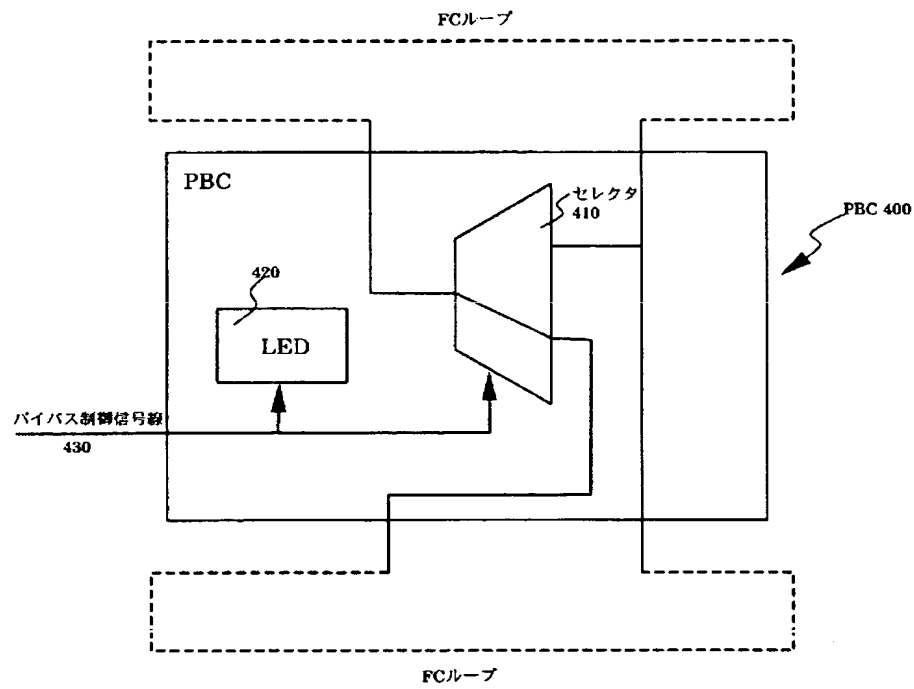
【図 3】

【図 3】

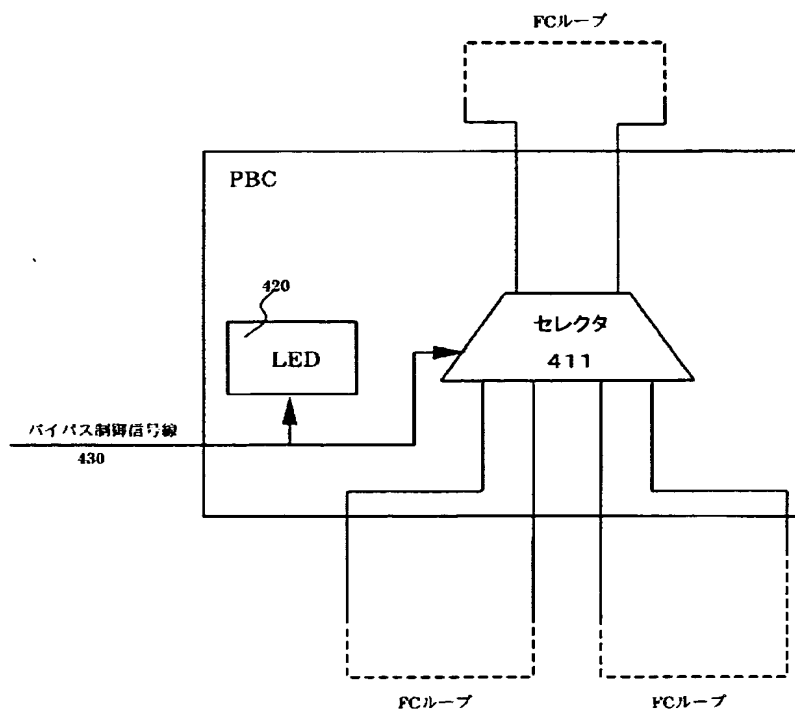


【図 4】

【図 4】

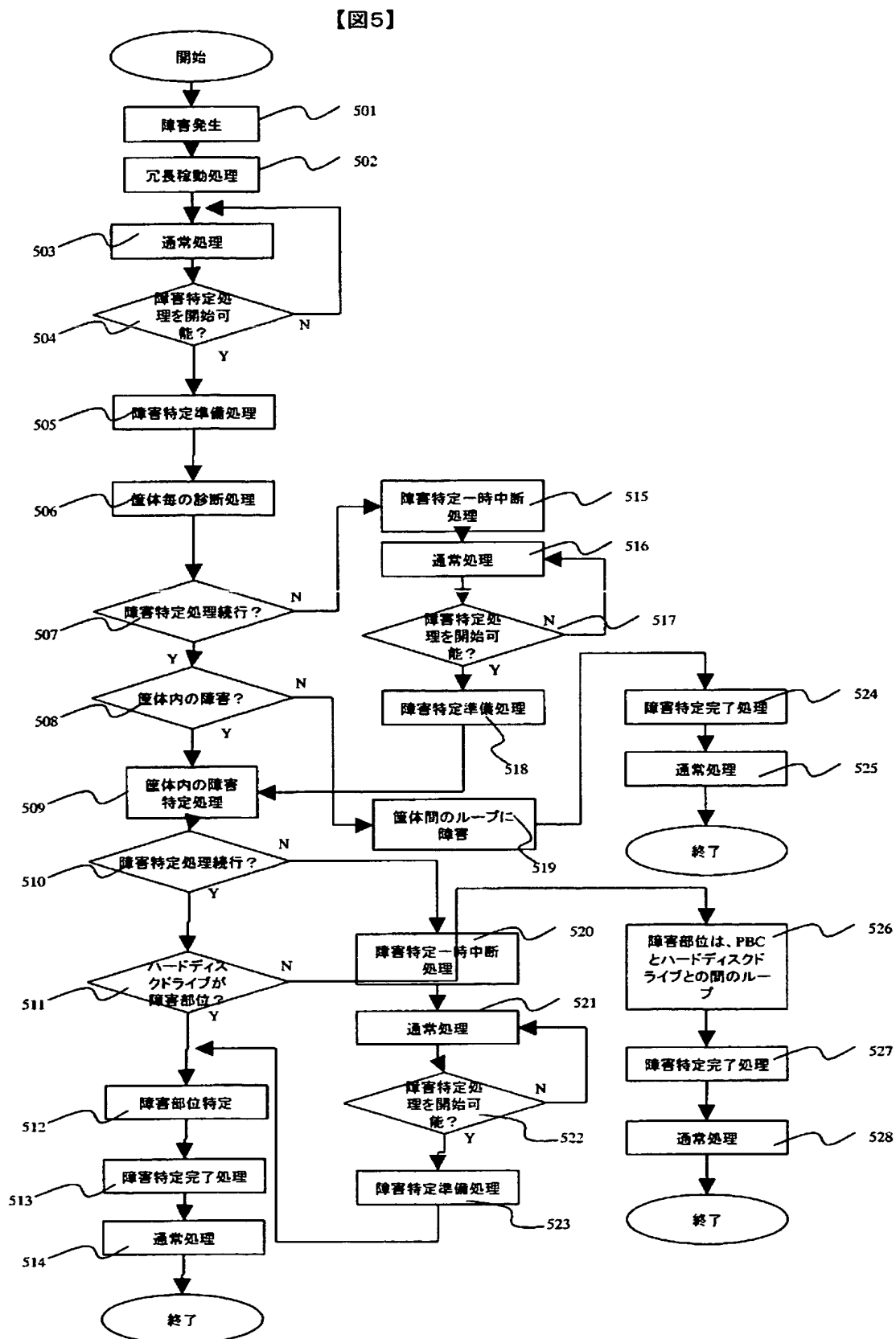


(a)



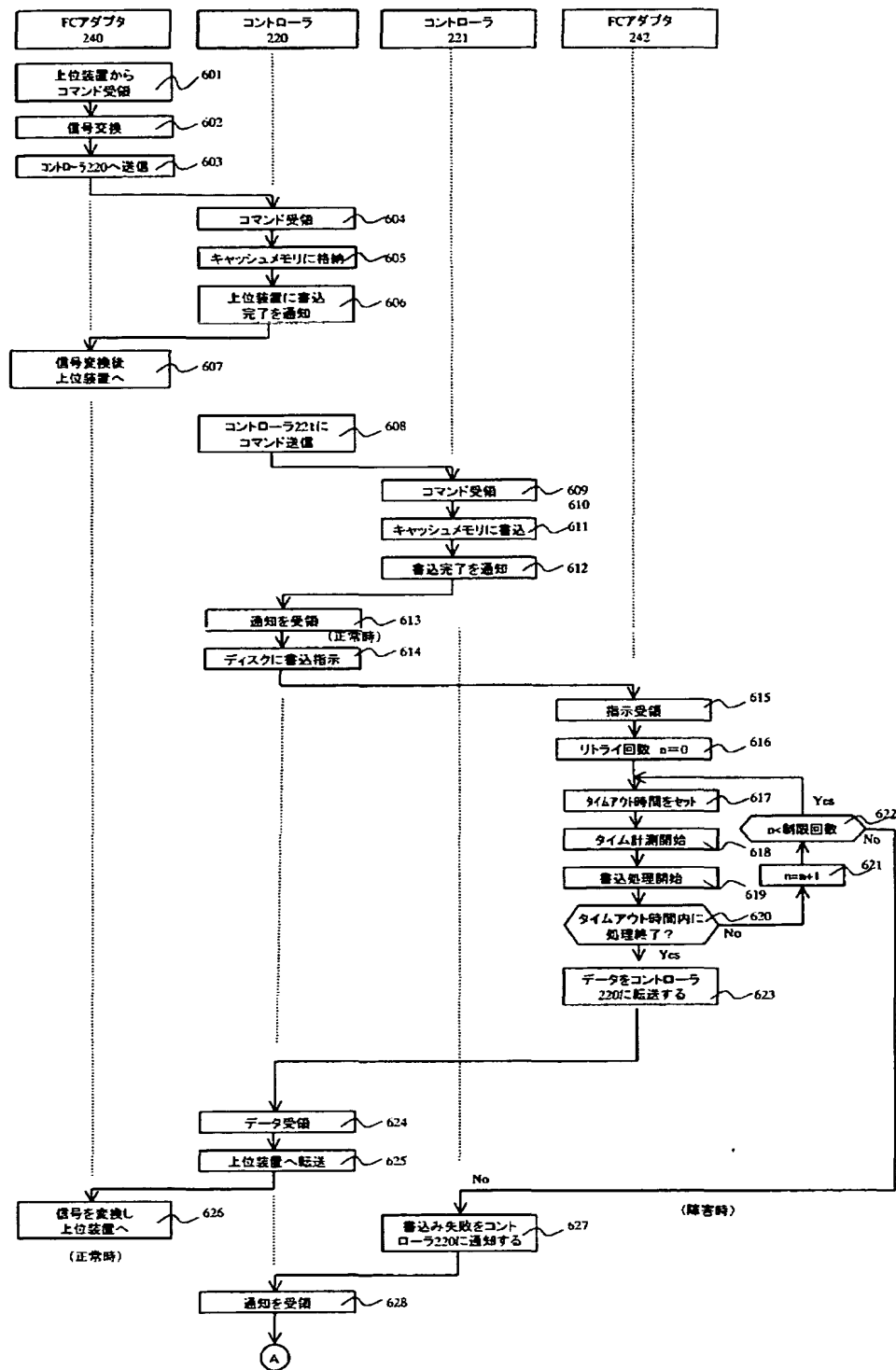
(b)

【図 5】



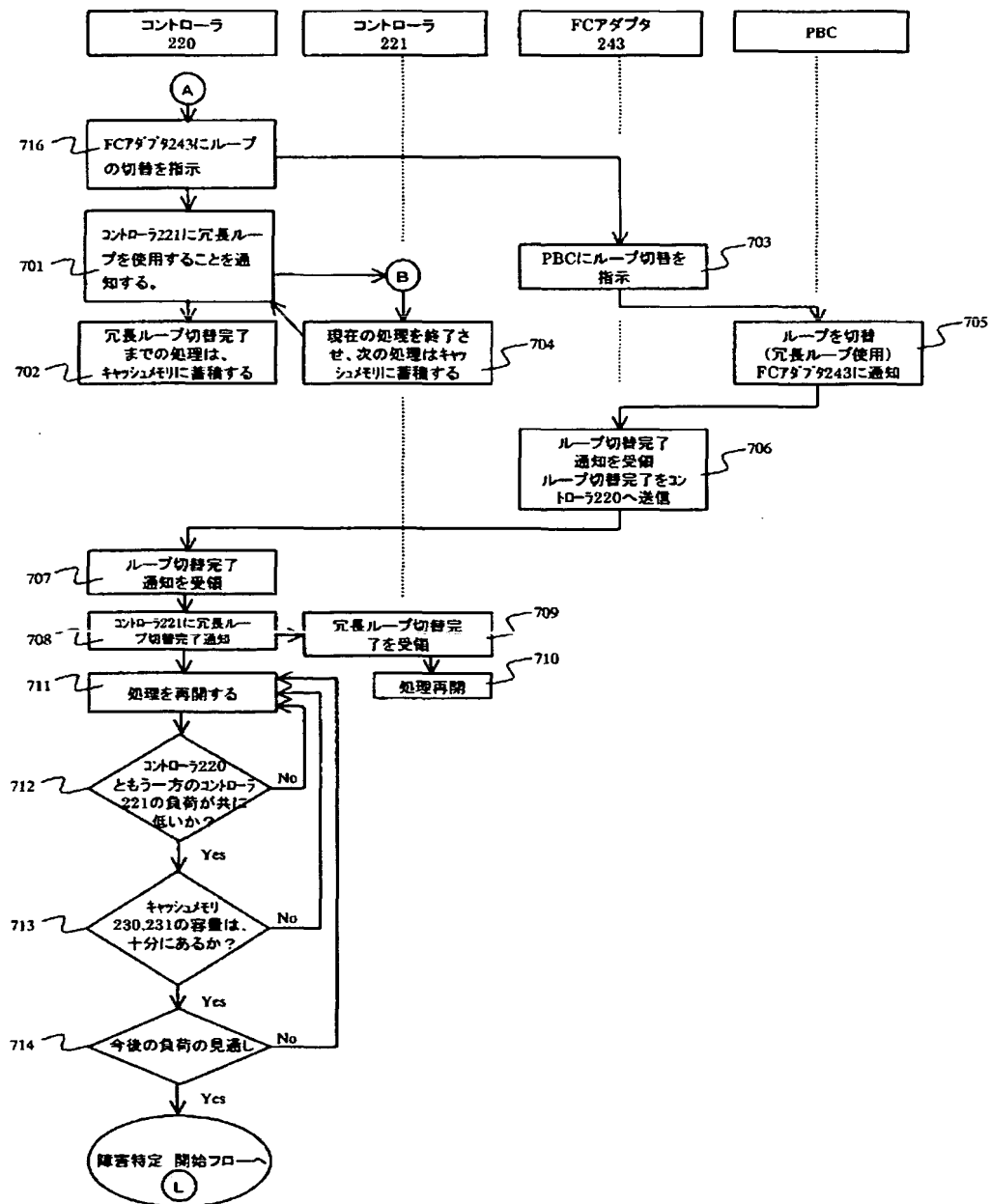
【図 6】

【図6】



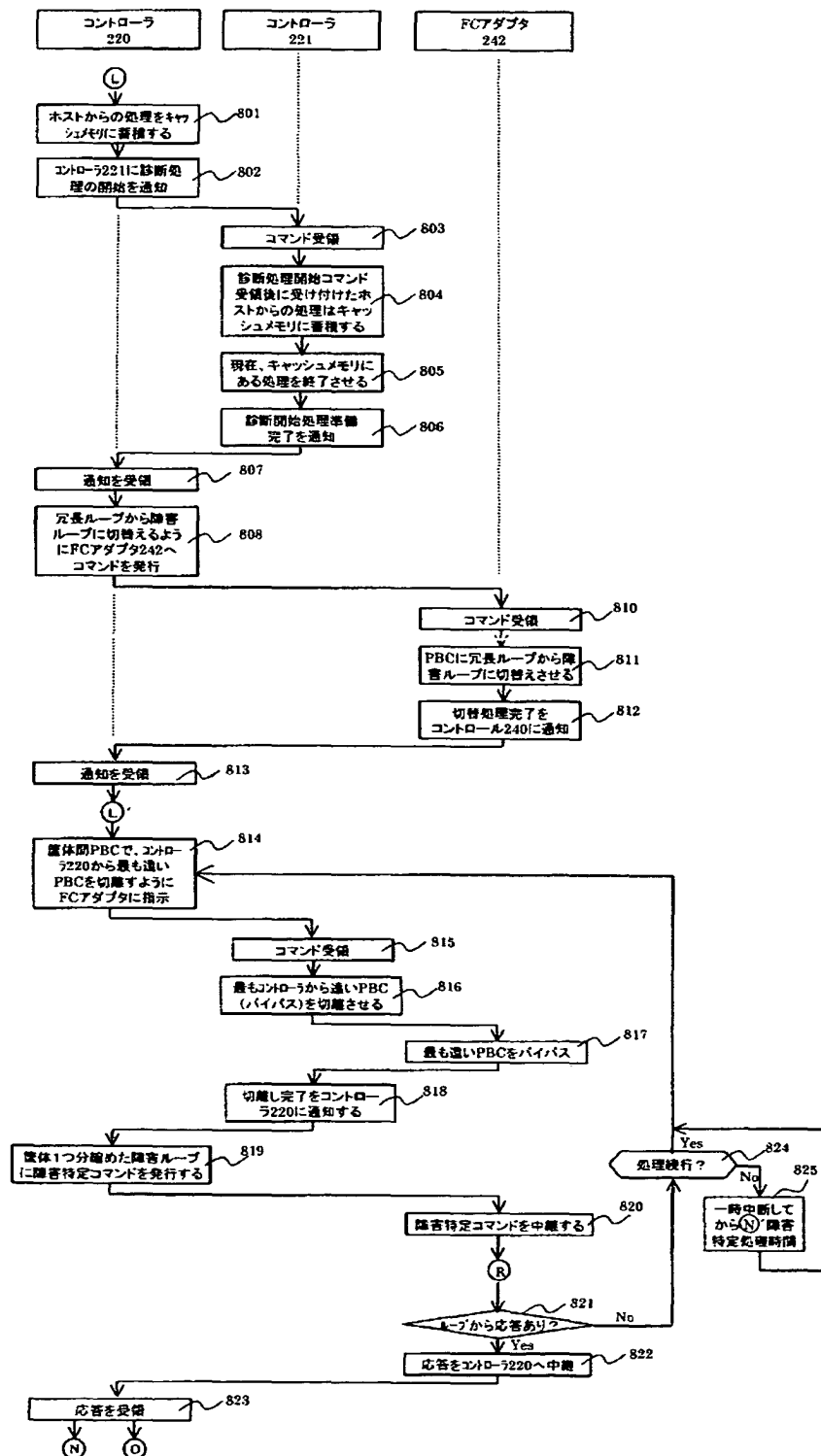
【図7】

【図7】



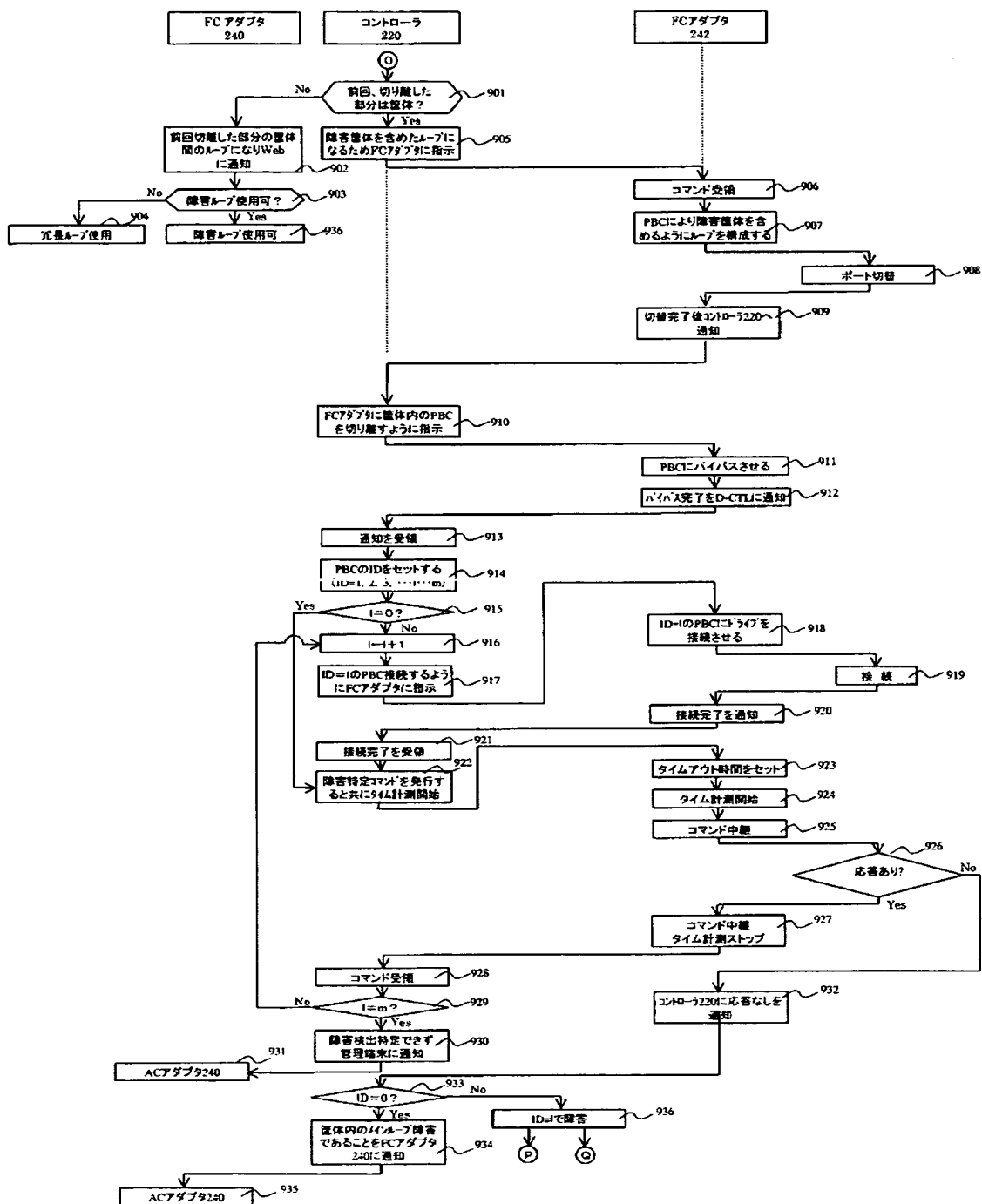
【図 8】

【圖8】

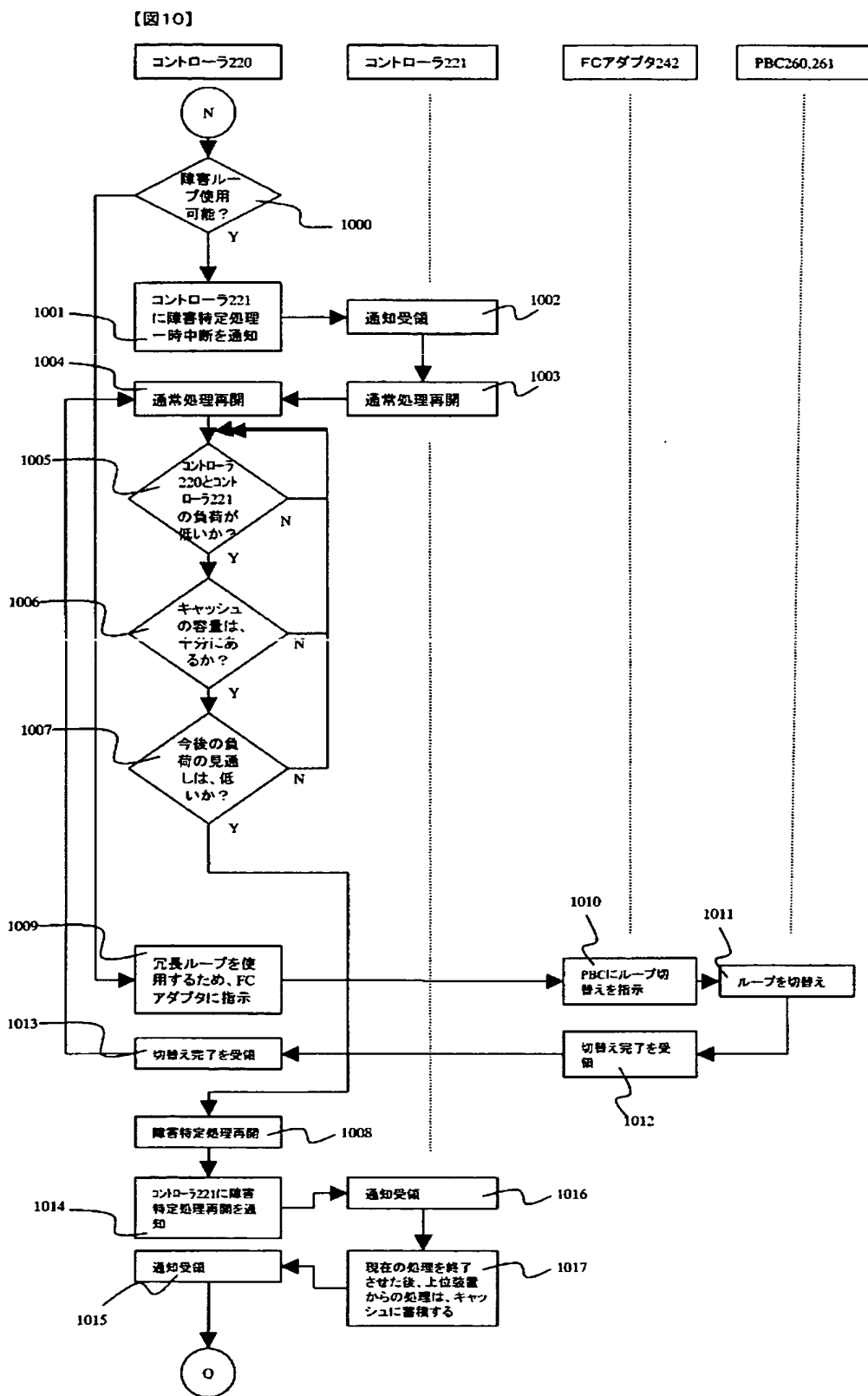


【図9】

【図9】

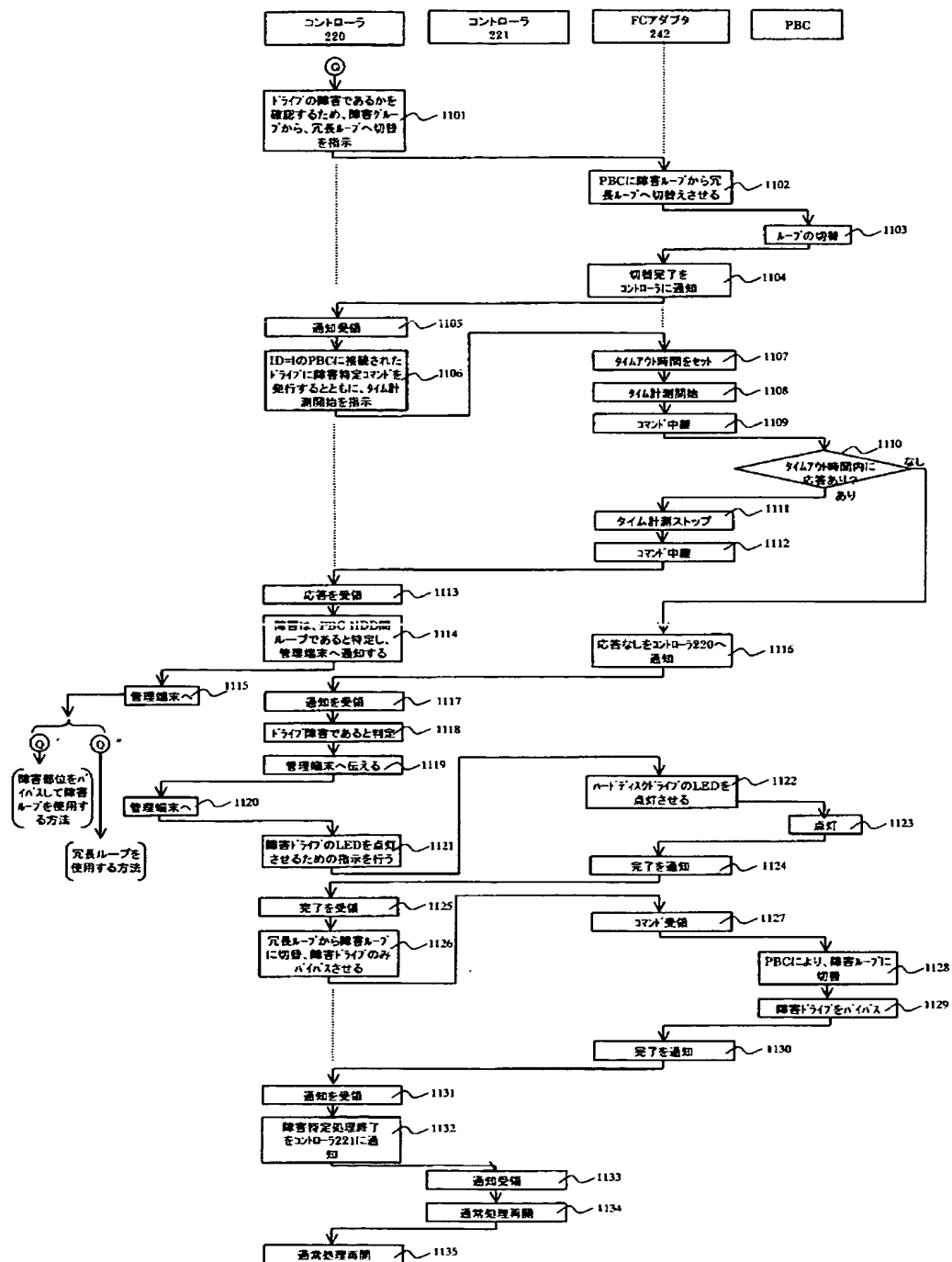


【図 10】



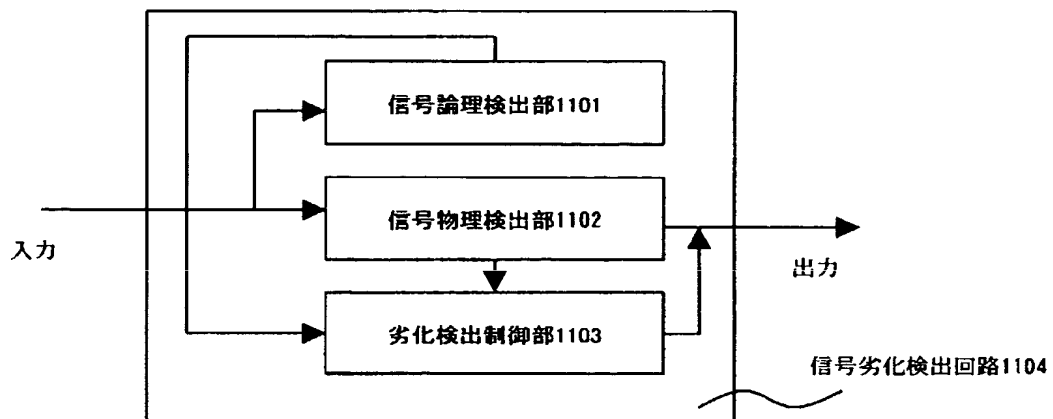
【図 11】

【図 11】



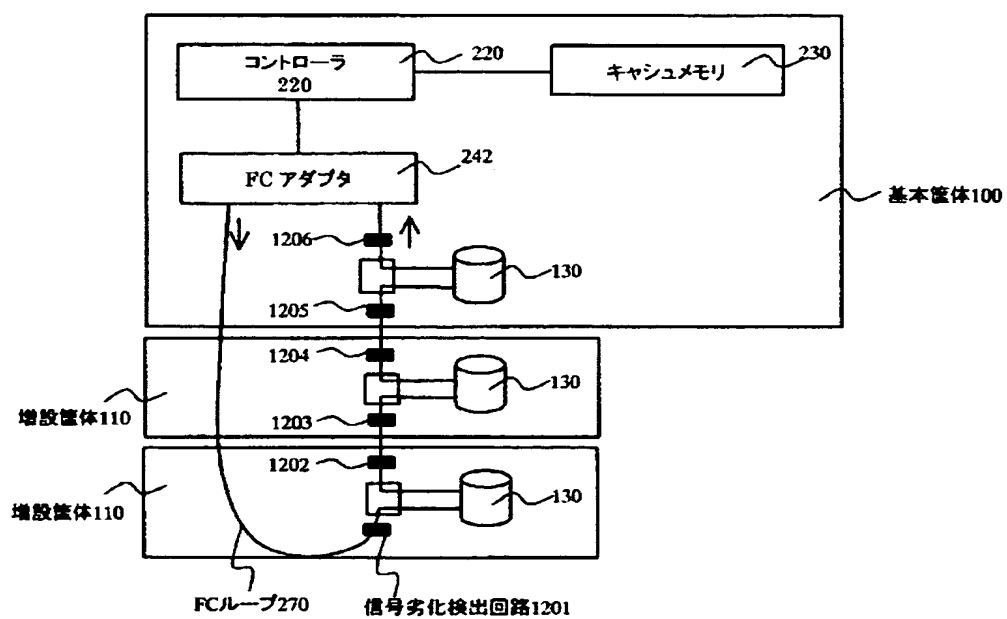
【図 12】

【図 12】

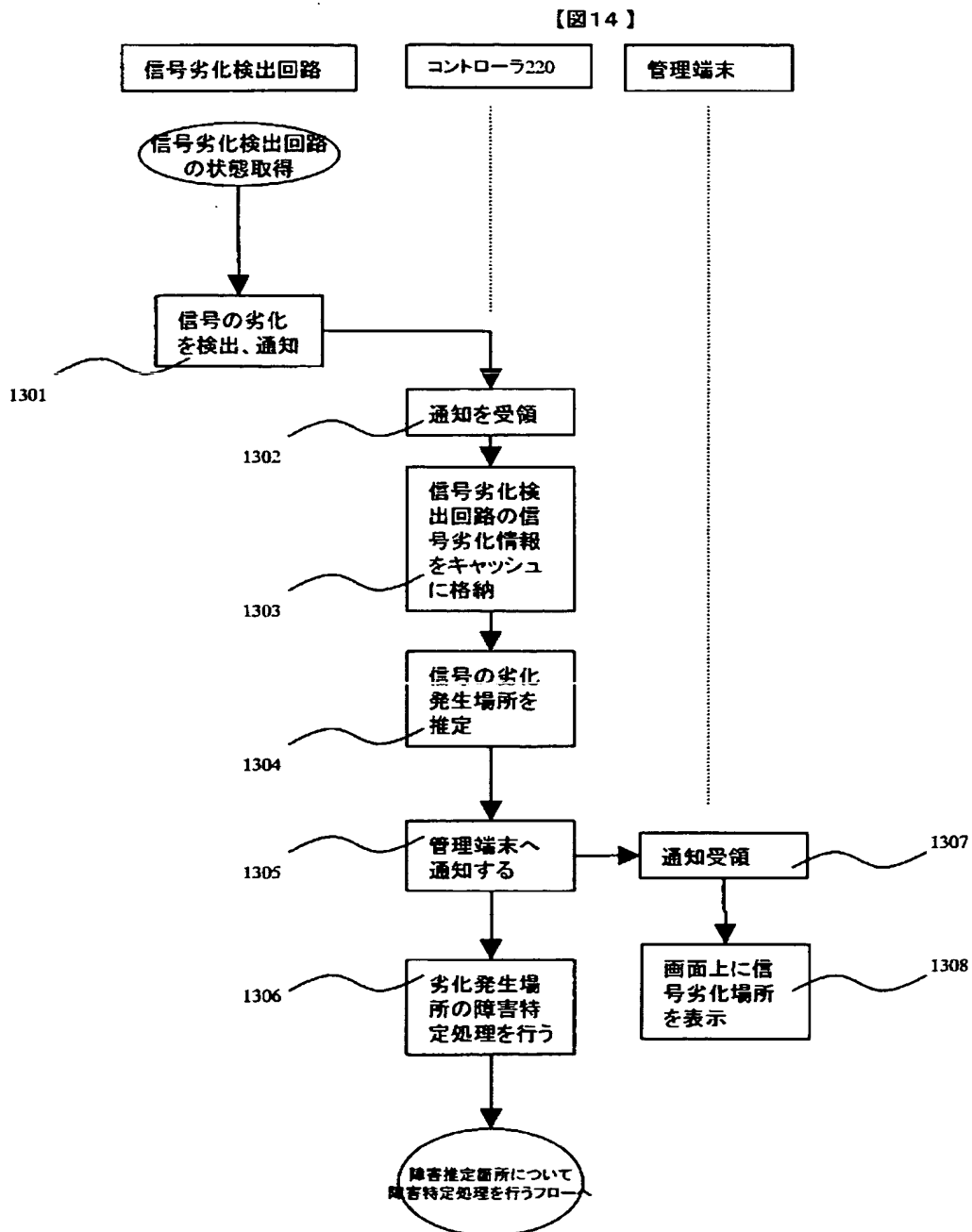


【図 13】

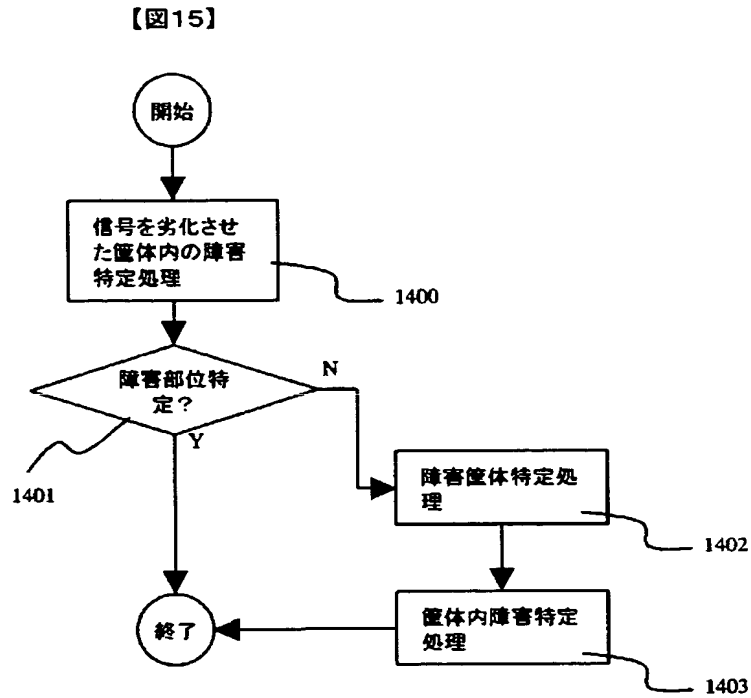
【図 13】



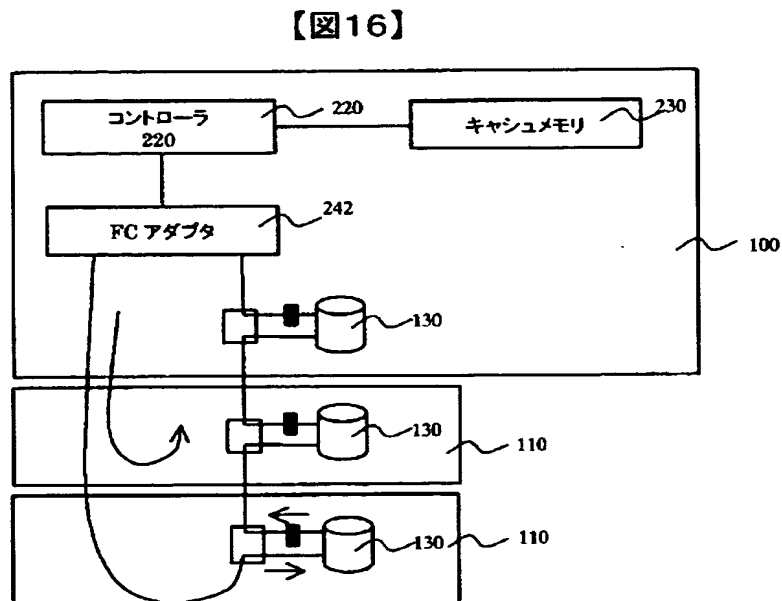
【図 14】



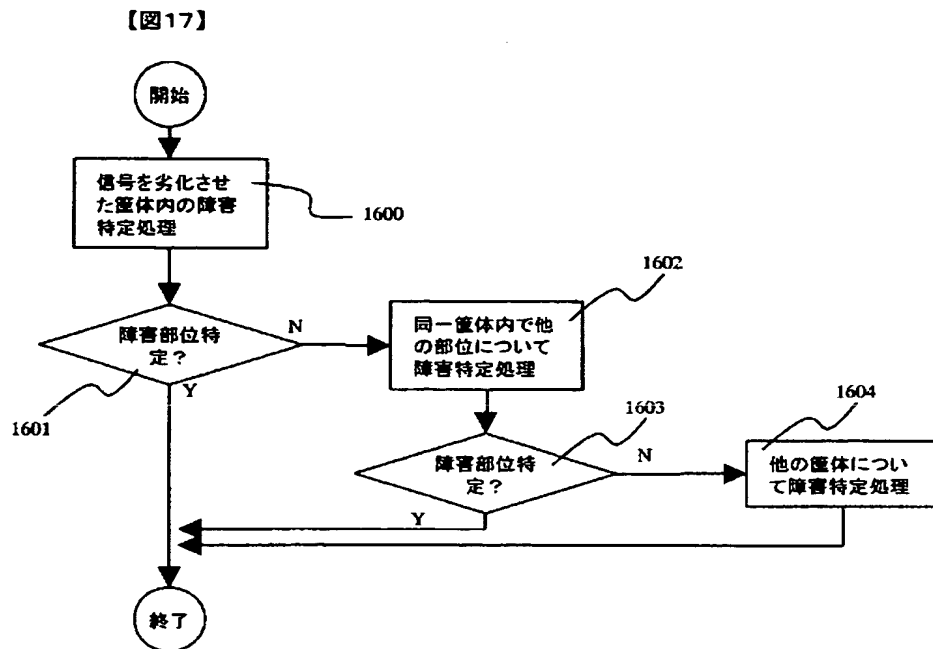
【図 15】



【図 16】

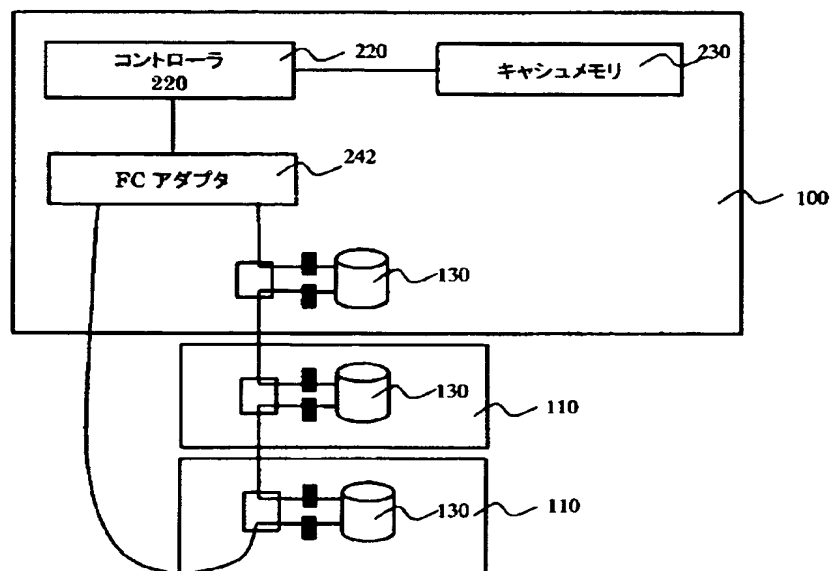


【図 17】

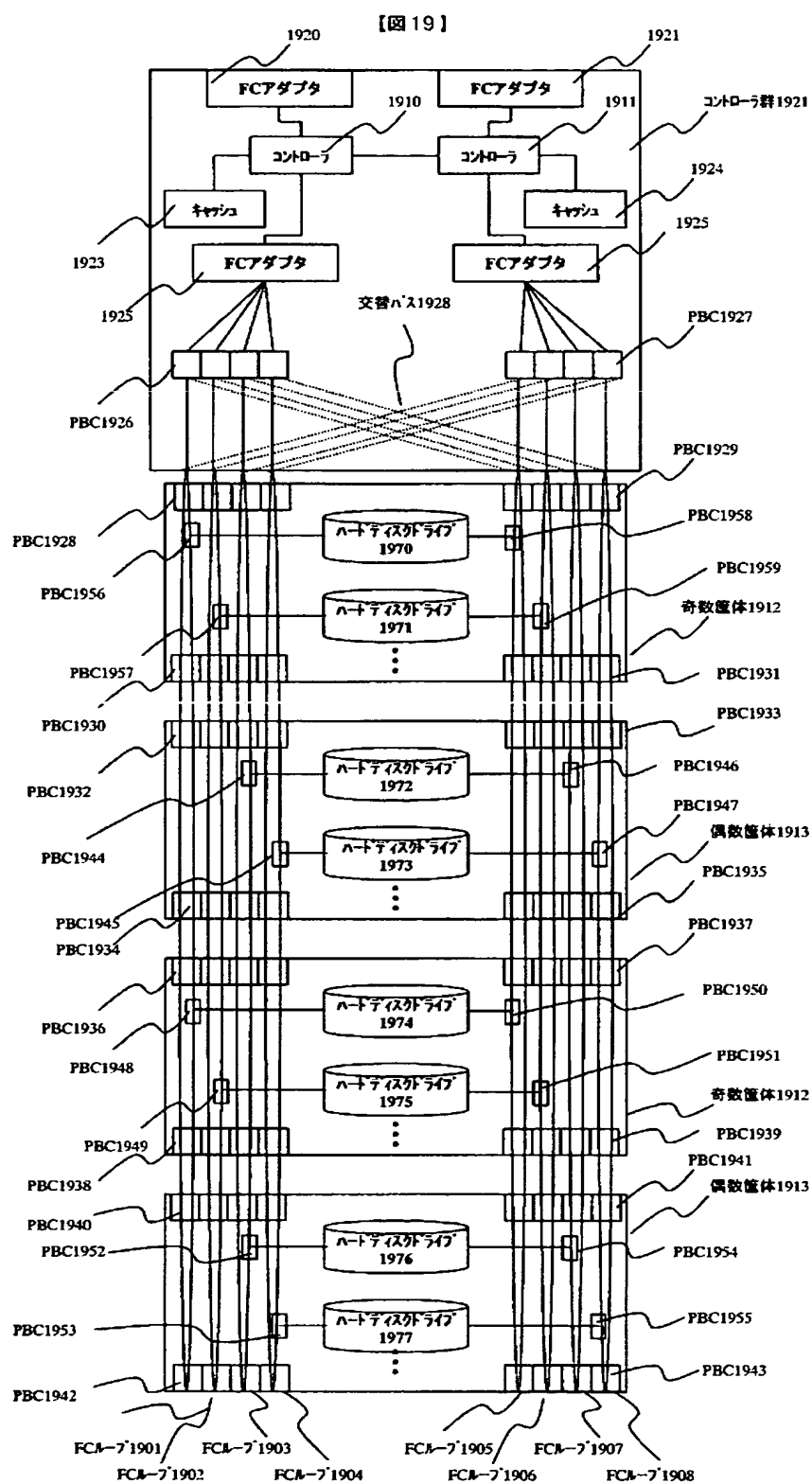


【図 18】

【図18】



【図 19】



【書類名】 要約書

【要約】

【課題】 記憶システムにおいて、通常処理を停止させることなく、障害部位を特定し、回復作業を行うことが可能な記憶システムを提供することにある

【解決手段】 少なくとも一つの記憶媒体と、記憶媒体を制御する少なくとも一つのコントローラと、記憶媒体及びコントローラをループ状に接続し、コントローラと記憶媒体との通信を相互に行うループ状通信手段とを含み、通常のリード・ライト処理を行いながら、障害部位を特定することができることを特徴とする。

【選択図】 図 3

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 1 9 5 4 5 0
受付番号	5 0 3 0 1 1 4 9 8 1 9
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 7 月 1 4 日

< 認定情報・付加情報 >

【提出日】 平成 15 年 7 月 11 日

特願 2 0 0 3 - 1 9 5 4 5 0

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1 . 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所